Precise Predictions for Top-quark Pair Production in Association with Multiple Jets

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Niccolò Moretti

aus

Italien

Promotionskomitee

Prof. Dr. Stefano Pozzorini (Vorsitz)

Prof. Dr. Florencia Canelli

Prof. Dr. Thomas Gehrmann

Prof. Dr. Massimiliano Grazzini

Zürich, 2016

Abstract

The discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC) has provided a spectacular confirmation of the Standard Model of particle physics and has opened a new era in the exploration of fundamental particles and forces at high-energy colliders.

In the coming years huge efforts will be devoted to the systematic investigation of the scalar and Yukawa sectors of the Standard Model. In particular, precise measurements of the Higgs couplings to force and matter particles and the comparison of such experimental data to theoretical predictions of comparable accuracy will provide a further decisive test of the Standard Model and possible evidence of new physics.

In this context, the top quark plays a particularly important role. Its very large mass, of the order of the electroweak scale, is widely believed to play a special role in the symmetry breaking mechanism and to be connected to physics Beyond the Standard Model. Moreover, top-production processes represent a dominant source of background in several Higgs-boson searches. The most striking example of this peculiar connection of top and Higgs physics is provided by the production of a top-quark pair in association of a Higgs boson ($t\bar{t}H$). Being the only Higgs-production mode that allows for a direct determination of the top-Yukawa coupling, $t\bar{t}H$ production can provide a unique test of the mechanism that generates the top-quark mass and is regarded as one of the most important processes at Run II of the LHC.

Due to the presence of overwhelming QCD backgrounds, searches for $t\bar{t}H$ production at hadron colliders are notoriously very challenging, and so far this crucial process could not be observed at the LHC. In the dominant channel, where the Higgs boson decays into a bottom–antibottom quark pair ($H \to b\bar{b}$), the $t\bar{t}H(b\bar{b})$ signal is obstructed by the production of a top-antitop pair in conjunction with multiple jets ($t\bar{t}$+multijets). In particular, $t\bar{t}$ production in association with two b-jets ($t\bar{t}b\bar{b}$) constitutes the dominant source of background.

The availability of accurate and realistic theoretical predictions for both signal and backgrounds is therefore of paramount importance in order to obtain significant sensitivity improvements in the $t\bar{t}H(b\bar{b})$ channel. This calls for next-to-leading order (NLO) calculations in perturbation theory, which ensure a significant reduction of theoretical uncertainties. Moreover, in order to achieve the degree of realism that is indispensable in order to apply

theoretical simulations to experimental analyses, NLO calculations must be matched to the parton shower (PS).

In the case of $t\bar{t}$+multijet production this task is particularly challenging. On the one hand, the calculation of the relevant Matrix elements at NLO QCD is very demanding, as it includes final states with four or more coloured particles. On the other hand, the multiscale nature of the problem, which involves various independent scales that range from the bottom mass, $m_b \sim 5\,\text{GeV}$, up to the invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}} \sim 500\,\text{GeV}$, leads to nontrivial issues in the context of NLO matching and in the assessment of theoretical uncertainties.

In this thesis we present predictions at NLO QCD for $t\bar{t}b\bar{b}$ and $t\bar{t}$+multijet production at the LHC. The relevant simulations have been performed with the Sherpa Monte Carlo event generator interfaced to the OpenLoops amplitude generator. In the case of $t\bar{t}b\bar{b}$ production, NLO matrix elements have been matched to the Sherpa Parton Shower including, for the first time, b-mass effects. This has allowed us to study, for the first time, a novel $t\bar{t}b\bar{b}$ production mechanism, which corresponds to kinematic configurations where one or two b–jets arise from a collinear $g \to b\bar{b}$ splitting and turned out to have an unexpectedly sizeable impact in the Higgs signal region. For $t\bar{t}$+multijet production we present results with up to three light jets at NLO accuracy. This calculation involves amplitudes with seven coloured external legs, two of which are heavy flavours, and presents a technical complexity of unprecedented level. To address the issue of scale choices and related uncertainties in the presence of multiple scales, we compared $t\bar{t}$+multijet results obtained with a standard scale choice at fixed order and the Multi-scale Improved NLO (MiNLO) procedure.

The simulations presented in this work represent state-of-the-art theory predictions for the $t\bar{t}H(b\bar{b})$ background and should be considered as a first step towards a unified precision simulation of $t\bar{t}$+multijet production including light and heavy-flavour jets. The accuracy reached in these simulations is a key prerequisite for observing for the first time the $t\bar{t}H(b\bar{b})$ signal, to measure the top and bottom Yukawa couplings, and to find evidence for possible new physics in the top-Higgs sector.

Zusammenfassung

Die Entdeckung des Higgs-Bosons im Jahr 2012 am Large Hadron Collider (LHC) lieferte eine spektakuläre Bestätigung des Standardmodells der Teilchenphysik und eröffnete ein neues Zeitalter der Erforschung der Elementarteilchen und fundamentalen Wechselwirkungen an Hochenergiebeschleunigern.

In den nächsten Jahren werden große Anstrengungen unternommen werden, den skalaren Sektor und den Yukawa-Sektor des Standardmodells zu untersuchen. Besonders die genaue Vermessung der Higgs-Kopplungen an Austausch- und Materieteilchen sowie der Vergleich solcher experimenteller Daten mit theoretischen Vorhersagen von vergleichbarer Präzision wird einen weiteren entscheidenden Test des Standardmodells sowie möglicherweise Hinweise auf Neue Physik liefern.

In diesem Zusammenhang spielt das Topquark eine besonders wichtige Rolle. Seiner sehr großen Masse, von der Größenordnung der elektroschwachen Skala, wird oft eine besondere Rolle beim Mechanismus der Spontanen Symmetriebrechung sowie eine Verbindung zu Physik jenseits des Standardmodells zugeschrieben. Außerdem stellen Topquarkproduktionsprozesse einen dominanten Hintergrund bei mehreren Suchen nach dem Higgs-Boson dar. Das auffälligste Beispiel für diese besondere Verbindung von Topquark- und Higgs-Physik ist die Produktion eines Topquarkpaares zusammen mit einem Higgs-Boson ($t\bar{t}H$). Da dies der einzige Higgs-Produktionskanal ist, der eine direkte Bestimmung der Top-Yukawa-Kopplung erlaubt, kann die $t\bar{t}H$-Produktion einen einzigartigen Test für den Mechanismus liefern, welcher die Topquarkmasse generiert, und wird als einer der wichtigsten Prozesse im RunII des LHC angesehen.

Wegen der überwältigenden Präsenz von QCD-Hintergrundprozessen sind Suchen nach der $t\bar{t}H$-Produktion an Hadronenbeschleunigern bekannt dafür sehr herausfordernd zu sein und tatsächlich konnte dieser entscheidende Prozess am LHC noch nicht beobachtet werden. Im dominanten Kanal, wo das Higgs-Boson in ein Bottom-Antibottom-Quarkpaar zerfällt ($H \to b\bar{b}$) wird das $t\bar{t}H(b\bar{b})$-Signal durch die Produktion eines Top-Antitop-Quarkpaares zusammen mit mehreren Jets ($t\bar{t}$+Multijets) verdeckt. Besonders die $t\bar{t}$-Produktion zusammen mit zwei b-Jets ($t\bar{t}b\bar{b}$) stellt einen dominanten Hintergrund dar.

Die Verfügbarkeit genauer und realistischer theoretischer Vorhersagen sowohl für das Signal als auch für die Hintergrundprozesse ist deshalb von überragender Bedeutung um signifikante Verbesserungen bei der Empfindlichkeit im $t\bar{t}H(b\bar{b})$-Kanal zu erzielen. Dies macht Berechnungen in nächstfüh-

render Ordnung (NLO) in Störungstheorie notwendig, welche eine deutliche Reduktion der theoretischen Unsicherheiten sicherstellen. Außerdem müssen NLO-Berechungen mit Partonenschauer (PS) kombiniert werden (matching), um den Realitätsgrad zu erreichen, der für die Anwendung theoretischer Simulationen in experimentellen Analysen unerlässlich ist.

Im Fall der $t\bar{t}$+Multijet-Produktion ist diese Aufgabe besonders herausfordernd. Einserseits ist die Berechnung der relevanten Matrixelemente in NLO QCD sehr fordernd, da hier Endzustände mit vier oder mehr Farbladung tragenden Teilchen auftreten. Andererseits führt das Auftreten verschiedener unabhängiger Energieskalen, welche von der Bottomquarkmasse $m_b \sim 5\,\text{GeV}$ bis hinauf zur invarianten Masse des $t\bar{t}$-Systems mit $m_{t\bar{t}} \sim 500\,\text{GeV}$ reichen, zu nicht-trivialen Problemen im Zusammenhang mit dem NLO matching und bei der Abschätzung der theoretischen Unsicherheiten.

In dieser Arbeit werden Vorhersagen in NLO QCD für die Produktion von $t\bar{t}b\bar{b}$ und $t\bar{t}$+Multijet am LHC vorgestellt. Die relevanten Simulationen wurden mit dem SHERPA Monte Carlo Ereignisgenerator in Verbindung mit dem OPENLOOPS Amplitudengenerator durchgeführt. Im Fall der $t\bar{t}b\bar{b}$-Produktion wurden die NLO-Matrixelemente an den Sherpa-Partonenschauer angepasst, wobei zum ersten Mal Bottomquarkmasseneffekte berücksichtigt wurden. Dies hat es uns erstmalig ermöglicht einen neuartigen Mechanismus der $t\bar{t}b\bar{b}$-Produktion zu untersuchen, der von kinematischen Konfigurationen herrührt, bei welchen ein oder zwei Bottomquarkjets von einer kollinearen Abspaltung der Form $g \to b\bar{b}$ stammen. Es stellte sich heraus, dass diese einen unerwartet großen Einfluss im Bereich des Higgs-Signals haben. Für die $t\bar{t}$+Multijet-Produktion werden Ergebnisse mit bis zu drei leichten Jets in NLO-Genauigkeit präsentiert. Diese Rechnung beinhaltet Amplituden mit sieben Farbladung tragenden äußeren Beinchen, von denen zwei schwere Quarkflavour sind, und ist von beispielloser technischer Komplexität.

Um das Problemfeld der Wahl von Energieskalen und verwandter Unsicherheiten in Anwesenheit mehrerer Skalen zu untersuchen, wurden die $t\bar{t}$+Multijet-Ergebnisse, welche mit einer Standardwahl für die Skalen in fester Ordnung berechnet wurden, mit den mittels der Multi-scale Improved NLO (MINLO) Methode berechneten Ergebnissen verglichen.

Die in dieser Arbeit vorgestellten Simulationen sind die aktuellsten Theorievorhersagen für den $t\bar{t}H(b\bar{b})$-Hintergrund und sollten als ein erster Schritt in Richtung einer einheitlichen Präzisionssimulation der $t\bar{t}$+Multijet-Produktion, einschließlich Jets von leichten und schweren Quarkflavourn, angesehen werden. Die in diesen Simulationen erreichte Genauigkeit ist eine Schlüsselvoraussetzung für die erstmalige Beobachtung des $t\bar{t}H(b\bar{b})$-Signals, für die Vermessung der Topquark- und Bottomquark-Yukawa-Kopplungen sowie für das Finden von Hinweisen auf mögliche Neue Physik im Topquark-Higgs-Sektor.

Contents

Introduction

The announcement of the discovery of the Higgs boson [1, 2] on the fourth of July 2012 by ATLAS and CMS collaborations has marked the beginning of a new era in particle physics. In the coming years, the Large Hadron Collider (LHC) will produce proton-proton collisions at the centre of mass energy of 13 TeV, and will then reach the design energy of 14 TeV. The huge amount of collected data will allow for precise measurement of a vast range of observables in the Higgs-boson sector. This will provide great opportunities to investigate still unexplored aspects of the spontaneous symmetry breaking mechanism, on which the Standard Model (SM) of particle physics [3–8] is based, and could also provide first hints of physics beyond the Standard Model (BSM).

The top quark plays a special role in connection with electroweak symmetry breaking, inasmuch as it is the only fermion with a mass of order of the electroweak scale and with a Yukawa coupling of the order of one. A direct measurement of the top Yukawa coupling at the LHC can be performed through Higgs production in association with a top-antitop pair ($t\bar{t}H$), but, due to the presence of severe QCD backgrounds, this channel has not been directly observed yet. The main decay channels of this process are $H \to \gamma\gamma$, $H \to WW, ZZ, \tau\tau$ with multi-lepton final states and $H \to b\bar{b}$. The $H \to \gamma\gamma$ decays provide the cleanest signature, but are very rare due to the very small branching ratio of $\mathcal{O}(10^{-3})$. The multi-leptons channels are characterised by a more significant but still modest production rate (with an effective branching ratio of the order of a few percent) and a decent signal to background ratio which ranges from 10% to 40% [9, 10], depending on the signal region. The dominant decay mode, $t\bar{t}H(b\bar{b})$, which would give simultaneous access to the top and bottom Yukawa couplings, is characterised by a branching ratio of about 60%. In the most recent combination of these channels, the ATLAS experiment has observed exclusion limits of 3 times the $t\bar{t}H$ SM cross section at 95% confidence level using about $13\,\mathrm{pb}^{-1}$ of data at 13 TeV [9]. This is

comparable to the level of sensitivity obtained in the combination of Run I data, where ATLAS and CMS have observed exclusion limits of, respectively, 3.1 and 4.5 times the SM cross section [11, 12].

In this PhD thesis we will focus on the dominant $t\bar{t}H(b\bar{b})$ channel and address the issue of its huge background contamination. In this particular channel, the ATLAS and CMS experiments have observed exclusion limits of respectively 3.4 and 4.2 times the SM cross section at 95% confidence level in Run I [13, 14], whereas using the first $13.3\,\mathrm{pb}^{-1}$ of Run II data, the ATLAS experiment has found a limit 4 times above the SM [9]. The low sensitivity in $t\bar{t}H(b\bar{b})$ searches is mainly due to systematic uncertainties, which are dominated by the theoretical uncertainties associated with the overwhelming QCD backgrounds. The dominant background consists of top-antitop production accompanied by multiple jet emissions ($t\bar{t}$+jets), in particular when a $t\bar{t}$ system is produced in association with a $b\bar{b}$ pair ($t\bar{t}b\bar{b}$), which corresponds to the irreducible background to the $t\bar{t}H(b\bar{b})$ signal. An additional difficulty arises from the presence of two additional b-quarks that result from $t \to Wb$ decays. This leads to combinatorial ambiguities in the selection of b-jets in the final state, which hamper a correct identification of the $H \to b\bar{b}$ mass peak, resulting in a dramatic dilution of the Higgs signal and, consequently, in a strong enhancement of its background contamination.

State-of-the-art selection techniques in conjunction with accurate and reliable theoretical predictions for $t\bar{t}H$ signal and backgrounds are therefore essential in order to achieve sufficient sensitivity in the $H \to b\bar{b}$ channel. Next-to-leading order (NLO) calculations in perturbation theory benefit of small intrinsic uncertainties, but they need to be matched to parton shower (PS) generators in order to obtain fully realistic simulations at particle level. In the last decade, several techniques to combine the accuracy of higher order calculations with the more physical description of final states provided by parton showers have been developed [15–26]. Such techniques have been implemented and automated in Monte Carlo (MC) event generators [25, 27, 28], which also support the modelling of the various nonperturbative aspects of particle collisions, namely the hadronisation, the underlying event and hadron decays. This renders MC theoretical simulations directly applicable to experimental analyses.

Theoretical predictions for the $t\bar{t}$+multijets background to $t\bar{t}H(b\bar{b})$ production at parton-level for top-quark pair production with two jets at NLO accuracy have been published in [29, 30]. These calculations can reduce perturbative uncertainties from 70–80% down to 15–20%. More recently, thanks also to the progress in terms of efficiency and speed of the employed automated tools, it has been possible to produce more realistic simulations at NLO of top-pair production plus multiple jet emissions, where

NLO matrix elements with up to one [31] and two [32] jets are merged and matched to the parton shower. Regarding the $t\bar{t}H(b\bar{b})$ irreducible background, $pp \to t\bar{t}b\bar{b}$, parton-level results at NLO accuracy [33–35] feature a 20–30% uncertainty. A NLO calculation matched to the parton shower using the POWHEG method [16, 17] has been published in [36], whereas in [37] the matching has been performed adopting the S-MC@NLO technique [18] taking also into account b-mass effects.

In this thesis we present precise and realistic theoretical predictions at next-to-leading order accuracy in QCD for the $t\bar{t}b\bar{b}$ and $t\bar{t}$+multijets backgrounds to $t\bar{t}H(b\bar{b})$. The manuscript is organised as follows. In Chapter 1 we review theoretical aspects of MC simulations. Starting from general concepts of NLO calculations at fixed order and parton shower resummation, we discuss state-of-the-art techniques for their combination. In particular we review the S-MC@NLO matching method, the MEPS@NLO merging technique and the MINLO algorithm [38] implemented in SHERPA. We also introduce the SHERPA+OPENLOOPS Monte Carlo framework, which has been employed throughout this work. In Chapter 2 we present a simulation of the process $pp \to t\bar{t}b\bar{b}$ at NLO accuracy published in [37], and complement it with more extensive unpublished studies. In Chapter 3 we present a very recent calculation of $t\bar{t}$+multijet production, including up to three jets at NLO accuracy [39]. In this context, to address the issue of scale choices and related uncertainties in the presence of multiple scales, we compare results obtained with a standard scale choice at fixed order and the MINLO procedure.

Finally we present the conclusions of this work. In the appendices the reader can find the explicit formulas used in the Sudakov form factors in the MINLO algorithm (Appendix A), a systematic comparison between Monte Carlo generators in the $t\bar{t}b\bar{b}$ production that will appear in the 4^{th} Yellow Report of the LHC Higgs cross section working group (Appendix B) and unpublished NLO predictions for $t\bar{t}$+multijet production (Appendix C).

Part I

Theoretical aspects and Tools

Chapter 1

Next-to-leading order matching and merging

In this chapter we discuss general methodological aspects of precision simulations at hadron colliders. In Section 1.1 we outline the essential features of next-to-leading order (NLO) calculations, with emphasis on the subtraction formalism, and we introduce Monte Carlo showers in Section 1.2. In Section 1.3 we discuss the matching and merging of NLO calculations and parton showers focusing on the S-MC@NLO method, whereas in Section 1.4 we discuss the MINLO method. Finally, in Section 1.5 we introduce the MC framework employed throughout this work, namely SHERPA+OPENLOOPS.

1.1 NLO calculations

A hadronic scattering event consists of a process where two hadrons h_1 and h_2 collide, generating a certain final state X. Hadrons are not fundamental particles, but rather bound states of fundamental constituents, the partons, that at high energies behave as free particles and take directly part in the interaction.

The typical hierarchy of energy scales, $\mathcal{O}(1)$ GeV for the confinement of partons inside hadrons and $\mathcal{O}(10^2$–$10^3)$ GeV for the scattering energy, allows one to factorize the problem in two different stages, and to consider them independently. More precisely, the cross section of a scattering process between two hadrons h_1 and h_2 into a final state X can be described by the following factorisation formula [40, 41]

$$\sigma = \sum_{i,j=q,\bar{q},g} \int_0^1 dx_i dx_j f_{i,h_1}(x_i,\mu_F) f_{j,h_2}(x_j,\mu_F) d\hat{\sigma}_{ij\to X}(\mu_F,\hat{s}). \tag{1.1}$$

The total cross section is a sum over all the possible partonic cross sections $d\hat{\sigma}_{ij\to X}$ that contribute to the X final state, where i and j are the partonic constituents, i.e. quarks and gluons, of h_1 and h_2, respectively. Each partonic cross section is convoluted with the Parton Distribution Functions (PDFs) $f_{i,h_1}(x_i,\mu_F)$ and $f_{j,h_2}(x_j,\mu_F)$, that describe the probability to find a parton $i(j)$ into the hadron $h_1(h_2)$ with momentum fraction $x_i(x_j)$ of the total momentum p. This factorized picture holds provided that the physics that describes the high-energy partonic interaction is independent from the low-energy physics that describes the hadronic structure and that is encoded in the PDFs. The partonic cross section, instead, depends on the particular process at hand, and can be written as

$$d\hat{\sigma}_{ij\to X} = \frac{1}{2\hat{s}} \int d\Phi_n |M_{ij\to X}(\Phi_n)|^2, \tag{1.2}$$

where $\hat{s} = x_i x_j s$ is the centre of mass energy squared of the partonic scattering and $|M_{ij\to X}|^2$ is the matrix element squared associated to the $ij \to X$ process, summed and averaged over the colour and helicity degrees of freedom associated with final- and initial-state partons, respectively. This is then integrated over the n-body (Born) phase space of the final state X

$$d\Phi_n = \prod_{l=1}^{n} \frac{d^3 q_l}{(2\pi)^3 2E_l} (2\pi)^4 \delta^{(4)} \left(p_i + p_j - \sum_{l=1}^{n} q_l \right). \tag{1.3}$$

The matrix element $M_{ij\to X}$ contains all the information of the interaction and is computable in quantum field theory as a perturbative expansion in the coupling constant of the theory. The first non trivial order of this expansion is usually associated with tree level Feynman diagrams, and it is known as Leading Order (LO) approximation. At LO, computations are rather straightforward, and the partonic cross section (1.2) reads

$$d\hat{\sigma}^{LO}_{ij\to X} = \frac{1}{2\hat{s}} \int d\Phi_n |M^{tree}_{ij\to X}(\Phi_n)|^2. \tag{1.4}$$

Next-to-Leading-Order (NLO) calculations are more involved. This order receives contributions with an additional power of the coupling constant, which can be classified into two categories: virtual amplitudes, where the contribution comes from the interference of tree level Feynman diagrams with *1-loop* diagrams, i.e., diagrams where a virtual particle is emitted and reabsorbed without appearing in the final state; or squared *real* amplitudes, where an extra particle is emitted. Loop diagrams contain a closed loop where the flowing momentum k is not fixed and must be integrated over,

while real-emission diagrams consist of tree-level diagrams with an extra external leg. The NLO partonic cross section can be written as

$$\hat{\sigma}^{NLO}_{ij\to X} = \frac{1}{2\hat{s}}\Bigg(\int d\Phi_n |M^{tree}_{ij\to X}(\Phi_n)|^2 + \int d\Phi_n |M^{virt}_{ij\to X}(\Phi_n)|^2 + \int d\Phi_{n+1} |M^{real}_{ij\to X}(\Phi_{n+1})|^2 \Bigg) \tag{1.5}$$

where the tree level amplitudes and the virtual amplitudes,

$$|M^{virt}_{ij\to X}|^2 = 2\mathrm{Re}\left[\left(M^{tree}_{ij\to X}\right)^* M^{1-loop}_{ij\to X} \right], \tag{1.6}$$

are integrated over the Φ_n phase space, whereas the real emission amplitudes are integrated over the Φ_{n+1} phase space. Virtual and real contributions are both divergent. There are two types of divergences that affect NLO calculations, namely infrared (IR) and ultraviolet (UV) divergences.

IR singularities come from kinematical configurations where, either in the loop integrals or in real emission phase space integrals, a momentum becomes soft (close to zero) and/or collinear to an external momentum. IR singularities arising after the integration of the real amplitudes over the Φ_{n+1} phase space in the soft/collinear regime, exactly cancel those appearing in the virtual amplitudes in the Born phase space. This is known as Kinoshita-Lee-Nauenberg theorem [42, 43]. More precisely, in hadronic collisions collinear divergences coming from initial-state collinear splittings do not cancel. However, such singularities can be reabsorbed into the PDFs through universal factorisation counterterms, known as *collinear counterterms*. This yields an additional term that must be added into formula (1.5) to get an IR finite result.

UV divergences instead arise from the integration of the loop integrals in the limit of infinite loop momentum. If the theory is renormalizable, these divergences can be cured by regularizing the integrals and renormalizing the theory. The former means to introduce a proper prescription for these integrals in order to turn the singularities into well-defined mathematical expressions. The most commonly employed regularisation scheme is *dimensional regularisation*. This approach consists of performing phase-space integrals in $d = 4 - 2\varepsilon$ dimensions, with ε being an infinitesimal quantity, and then analytically continuing the results to the standard four dimensions. In the limit $\varepsilon \to 0$, divergences arise as $1/\varepsilon$ poles. The renormalisation procedure consists of a reinterpretation of the parameters of the theory (coupling and masses) in terms of "bare" and "renormalized" quantities. Such procedure introduce new tree-level effective interactions in the Lagrangian (so called

counter-terms) whose contributions exactly cancel the singularities arising from the the virtual amplitudes. The renormalisation procedure leads also to a spurious dependence of the scattering amplitude on an arbitrary *renormalisation scale* μ_R. However this dependence enters formally at the next order, i.e. only at the next-to-next-to leading order in an NLO calculation, and vanishes when the perturbative series is re-summed at all orders.

In summary, the NLO partonic cross section (1.5) can be written as

$$\hat{\sigma}^{NLO}_{ij\to X} = \frac{1}{2\hat{s}}\Bigg(\int d\Phi_n |M^{tree}_{ij\to X}(\Phi_n;\mu_R)|^2 \quad + \int d\Phi_n |M^{virt}_{ij\to X}(\Phi_n;\mu_R)|^2 + \int d\Phi_n |M^{tree}_{ij\to X}|^2 \otimes \mathcal{C}(\Phi_n;\mu_R,\mu_F) \quad + \int d\Phi_{n+1} |M^{real}_{ij\to X}(\Phi_{n+1};\mu_R)|^2 \Bigg), \tag{1.7}$$

where $\mathcal{C}$ is the additional PDF counterterm mentioned above, and the μ_R and μ_F dependences have been made explicit.

Despite all the ingredients might be available, this computation involves $3n$– and $3(n+1)$–dimensional phase-space integrals, whose analytical solution, in the case of multi-particle final states, is not known. Moreover, in order to produce useful predictions for experimental analyses, one should also integrate (1.4) with realistic cuts that correspond to the geometry of the detector and implement also experimental selection criteria. In general, this task cannot be addressed with analytic integration and requires a numerical approach.

On the one side, Monte Carlo (MC) methods are notoriously convenient for the calculation of multi-dimensional integrals. On the other side, each integrand must be finite in order to be numerically calculated, i.e., the cancellation of the divergences must occur locally at the integrand level. Various techniques have been elaborated to this end. They can be categorized in two main branches: phase space slicing [44, 45] and infrared subtraction methods [46, 47]. Throughout this work, we will focus on subtraction methods in the context of NLO QCD calculations. These methods are based on the observation that QCD IR singularities exhibit a universal structure. Hence they can be handled in a process independent way. In proximity of the singularities, squared real emission amplitudes factorize in Born-like amplitudes times universal splitting kernels which fully contain their IR singularities. The basic concept is then to locally remove such configurations from the real phase space and then re-add them in the Born phase space after analytical integration of the extra unresolved parton. This integration can be performed once and for all due to the universal structure and the relative simplicity of the splitting kernels. The subtracted integrands are separately finite, and thus suitable for MC evaluation.

Let us reexpress (1.1) in a more compact way. The differential NLO cross section for a general $2 \to n$ process in the subtraction formalism can be written as

$$d\sigma_{NLO} = d\Phi_B \left[B(\Phi_B) + V(\Phi_B) + I(\Phi_B)\right] + d\Phi_R \left[R(\Phi_R) - S(\Phi_R)\right], \quad (1.8)$$

where Φ_B and Φ_R represent the Born and real phase space, respectively, while B, V and I involve the contributions from Born matrix elements, virtual matrix elements and integrated subtraction terms. Finally, R and S are the real emission contributions and the corresponding subtraction terms. Flux factors, parton luminosities and symmetry factors are implicitly included in the phase space measures $d\Phi_B$ and $d\Phi_R$. As discussed above, the real subtraction terms S are required to match the singularities of the real amplitudes R in the soft-collinear limits. In this way, IR singularities are locally removed from the real phase space Φ_R and re-added in the Born phase space Φ_B, after analytic integration over the radiated parton, in the form of I-terms,

$$I(\Phi_B) = \int_1 d\Phi_1 S(\Phi_R). \quad (1.9)$$

Here the real phase space is factorised into a Born phase space and a one-particle emission phase space, $\Phi_R = \Phi_R(\Phi_B, \Phi_1)$. We will call Φ_B the underlying Born configuration of a real-emission event $\Phi_R(\Phi_B, \Phi_1)$. This procedure guarantees that each of the square brackets in (1.8) is finite[1]. A general observable $\mathcal{O}$ can be calculated by integrating (1.8) in the respective phase space, that is

$$\begin{aligned}\langle \mathcal{O} \rangle = &\int_n d\Phi_B \left[B(\Phi_B) + V(\Phi_B) + I(\Phi_B)\right] \mathcal{O}_n(\Phi_B) \\ &+ \int_{n+1} d\Phi_R \left[R(\Phi_R)\mathcal{O}_{n+1}(\Phi_R) - S(\Phi_B, \Phi_1)\mathcal{O}_n(\Phi_B)\right], \end{aligned} \quad (1.10)$$

provided that the observable is infrared safe, which means that, in the soft and collinear limits it must fulfil $\mathcal{O}_{n+1}(\Phi_R(\Phi_B, \Phi_1)) \to \mathcal{O}_n(\Phi_B)$.

In the most commonly employed subtraction methods [46,47], subtraction terms can be written as a convolution of the Born matrix elements with suitable operators K, such that

$$\begin{aligned} d\Phi_R S(\Phi_R) &= d\Phi_B d\Phi_1 \left[B(\Phi_B) \otimes K(\Phi_1)\right] \\ I(\Phi_B) &= \int_1 d\Phi_1 \left[B(\Phi_B) \otimes K(\Phi_1)\right]. \end{aligned} \quad (1.11)$$

[1] Actually, this formalism guarantees that both $V + I$ and $R - S$ are *locally* finite.

As long as the subtraction kernels K yield the same singularity structure of R, their definition is not unique, but the results obtained integrating (1.10) do not depend on the particular choice.

NLO calculations provide more accurate predictions than LO ones. In particular, NLO predictions feature a reduced sensitivity to the choice of renormalisation and factorisation scales. Moreover, at NLO, variations of μ_R and μ_F can be used as a meaningful tool to estimate theoretical uncertainties associated to unknown higher-order effects beyond NLO. Fixed order (FO) results, however, are based on an idealized representation of the scattering events, where the interaction is described by matrix elements (MEs) for the scattering of partons (quarks and gluons) at high energies. But this differs from what is actually measured in high energy collision experiments, where rather hadrons are detected in the calorimeters. The latter are (meta) stable particles that arise from colourless combinations of partons at low energy scales of $\mathcal{O}(1)$ GeV. In order to improve fixed-order results towards a more realistic hadron-level description of scattering events, we need hence to include the modelling of the evolution of partons from their production at high energy down to hadronisation energies. Such evolution can be implemented via parton shower (PS) generators. While parton showers can be easily applied to LO MEs, the combination of NLO predictions with parton showers is not trivial. The procedure to properly combine these approaches, known as "matching", is discussed in Section 1.3, while Section 1.2 reviews the essential aspects of parton shower generators.

1.2 Parton Showers

The description of hard interactions is provided by FO calculations, where the transition probability between an initial state i to a generic final state f (usually partons in case of QCD interactions) is expressed in terms of MEs using a perturbative approach. The description of how such partons evolve from the hard scattering down to hadronisation energies is provided by parton showers. In this evolution, partons progressively loose energy through multiple QCD emissions, similarly as in the case of QED bremsstrahlung, where positrons and electrons emit multiple photons. These multiple emissions turn out to be enhanced in phase space regions where the emitted partons are soft or collinear. However, fixed order amplitudes are divergent in these limits, thus a different approach is needed to properly describe parton evolution. Parton showers are based on the universal properties of factorisation of collinear singularities in QCD. Thus they can be employed to improve FO predictions in a process-independent manner. The key aspects of

parton-shower methods are outlined in this section. For a detailed discussion of the topic see [48, 49] and references therein.

Let us consider the emission of an extra collinear parton from a n-parton final state. In this limit the additional emission factorizes and the $2 \to n+1$ cross section takes the form

$$d\sigma_{n+1}(\Phi_{n+1}) \to \sum_{i,j=q,\bar{q},g} d\sigma_n \frac{\alpha_S}{2\pi} \frac{dt}{t} dz P_{ji}(z,\phi) \frac{d\phi}{2\pi}, \tag{1.12}$$

where the sum runs over all the external flavours i contributing to the $d\sigma_n$ cross section and all their possible collinear emitted partons j, while t is a variable which parametrizes such collinearity, such as the p_T^2 of the emitted parton relative to the emitter or the emission squared angle θ^2. P_{ji} is the DGLAP splitting function which encodes the collinear limit of the splitting $i \to jk$ with momentum fraction z and azimuthal angle ϕ. It should be emphasized that this result is completely general, i.e. it does not depend on the explicit form of the parametrisation variable t, and that different parametrisations lead to different finite terms accompanying the divergences. The interpretation of (1.12) is as follow: in the collinear limit, each parton acts independently from the others. This picture can thus be iterated for multiple emissions, taking into account that the validity of the factorisation formula and the appearance of the dt/t singularity in (1.12) are restricted to kinematic regions where t is the smallest relative transverse momentum in the $2 \to n+1$ process. This perfectly suits a Markov chain algorithm, and results also relatively straightforward to implement in a Monte Carlo generator. The limit $t = 0$ is associated with the collinear divergence and thus it should be regularized with a cut off t_{IR}. Emissions with $t > t_{IR}$ are said to be resolvable and are associated to physical (finite) probability distributions, whereas unresolvable emissions ($t < t_{IR}$) lead to divergences that will be cancelled by the respective virtual contribution. The dt/t logarithmic enhancements described by the factorisation formula (1.12) are restricted to the phase space regions with ordered emissions, i.e. if the n-th parton has been emitted at a certain scale Q^2, the following parton is effectively emitted at a scale t lower than Q^2. Let us now consider the probability that a resolvable emission occurred at a scale t. This is equivalent to studying the probability that no resolvable emission occurred between the scales Q^2 and t for the parton flavour i, which we define $\Delta_i(t, Q^2)$. The variation of the no emission probability between t and $t + dt$ corresponds to the probability that no emission occurred above t times the variation of the splitting (kernel) function due to the infinitesimal variation dt, i.e.

$$\frac{d}{dt}\Delta_i(t,Q^2) = \Delta_i(t,Q^2) \frac{\alpha_S}{2\pi t} \int_{z_0}^{1-z_0} dz P_{ji}(z), \tag{1.13}$$

where $P_{ji}(z)$ is the azimuthal averaged splitting function related to the splitting $i \to jk$, and $z_0 = z_0(t, t_{IR})$ (or $z_0(t, t')$, depending on the parametrisation) is the cutoff on z, which regularizes the IR limits $z = 0$ and $z = 1$ [2]. This differential equation can be integrated with the obvious boundary condition $\Delta_i(Q^2, Q^2) = 1$ (that is, the probability to not emit any parton between two identical scales is one) and gives the explicit solution

$$\Delta_i(t, t') = \exp\left\{ -\int_t^{t'} dt \frac{\alpha_S(t)}{2\pi t} \int_{z_0}^{1-z_0} dz P_{ji}(z) \right\}. \tag{1.14}$$

This function is known as *Sudakov form factor*. It is worth to notice that we obtained the above equation using unitarity arguments, that is, the probability to emit a parton between two scales plus the probability to not emit any parton is equal to one. Since the probability to not have any resolvable emission, order by order, is finite, and since divergences cancel at each order thanks to the virtual contribution, this approach indirectly introduces (partial) virtual corrections from tree level relations via unitarity arguments. The implementation of PS algorithms into MC event generators can be done in the following way. Given a starting scale Q, in order to generate a random variable t having a distribution function described by (1.14), one generates a (pseudo)random number $r_1 \in [0, 1]$ and solves the equation

$$\Delta(t_1, Q) = r_1.$$

The following emission $t_2 < t_1$ can occur from any of the external legs when

$$\Delta(t_2, t_1) = r_2,$$

and so on, until the emission scale falls below the cutoff t_{IR}. Below such scale, the coupling constant becomes large, perturbation theory does not hold anymore and we enter the regime of confinement . In this regime one must resort on *hadronisation models*.

Equation (1.14) has been obtained studying QCD evolution effects on partons that outgo from the hard scattering. For this reason we talk about final-state radiation (FSR). Likewise, we expect incoming partons to give rise to initial-state radiation (ISR) in a similar manner. Despite the physics is almost the same, the production mechanism described above would be extremely inefficient in the ISR case: most of the radiation generated by a parton shower starting at the proton scale and evolved up to the hard factorisation scale would be at low energy and low virtuality, and thus only very few

[2] In the compact notation employed here and in the following, the sum over all the possible splittings $i \to jk$ of the parton of type i is understood.

configurations would be compatible with the kinematic of the incoming partons needed as a input for the hard process. For this reason, parton shower algorithms are based on a backward evolution [50] for the simulation of ISR, which means that one generates the probability that a parton with a given momentum fraction and a given evolution scale originates from another parton with higher momentum fraction and lower evolution scale. This is then iterated until the evolution scale reaches the cutoff, where eventually non-perturbative models describe the extraction of the parton from the incoming hadron and the evolution of subsequent remnants. The Sudakov form factor for non-emission probability in case of ISR reads

$$\Delta_i(t,t';x) = \exp\left\{ -\int_t^{t'} \frac{\alpha_S(t)}{2\pi}\frac{dt}{t}dz\frac{x/z f_j(x/z,t)}{x f_i(x,t)}P_{ij}(z,\phi)\frac{d\phi}{2\pi}\right\}, \qquad (1.15)$$

where each emission is weighted by the proper parton distribution function ratio which accounts for the probability that its "current" momentum fraction x has been generated by a parent parton with higher momentum fraction $x' = x/z$.

The discussion made so far was based on collinear parton emissions. In this case the factorisation occurs at cross section level because one emitted parton can be collinear to one external parton at a time. The situation is different in case of the soft limit, where soft gluon emissions give rise to singularities as well, but their production occurs coherently from different partons and thus they should be treated at amplitude level. This coherence effect seems to spoil the factorized and recursive picture depicted so far. However, from explicit calculations it turns out that a coherent emission of wide-angle soft gluons occurring at a certain step by a parton pair can be treated *as if* it was produced by their "parent" parton before the more collinear splitting occurred . It should be emphasized that this is only the result of quantum interference effects, and that it does not correspond to any Feynman diagram where a gluon is emitted by an internal line. This result leads to *angular-ordered* showers, such as HERWIG [51], where the ordering variable is the opening angle θ itself. Another alternative is given by *dipole* showers. These methods are based on the decomposition of amplitudes in colour flows, i.e. in sets of colour lines connecting quarks and gluons that, in the soft region and in the limit of large number of colour degrees of freedom (N_c), emit independently, in a way that automatically accounts for soft gluon coherence. Each configuration is associated with a colour flow on a probabilistic manner, which then is iteratively showered according to formulas that are closely analogous to standard shower equations. Instead of sequential splittings of one parton into two, dipole showers can be considered as $2 \to 3$ splittings, where the additional coloured parton (the *spectator*) is not directly involved

in the splitting but allows for energy-momentum conservation at each emission step. Examples of dipole showers are ARIADNE [52] and the CSS [53] Catani–Seymour shower implemented in SHERPA.

As already pointed out, the factorisation encoded in (1.12) is exact in the collinear limit, and its structure does not depend on the explicit form of the parametrisation. Let us now generalize and re-write (1.14) in a more compact form,

$$\Delta(t,t') = \exp\left\{-\int_t^{t'} d\Phi_1 \frac{\alpha_S}{2\pi}\mathcal{K}(\Phi_1)\right\}, \tag{1.16}$$

where $t = t(\Phi_1)$ is the variable that parametrizes the degree of collinearity, and $\mathcal{K}(\Phi_1)$ is a splitting kernel which encodes the correct IR limits, i.e., which satisfies

$$\mathcal{K}(\Phi_1) \underset{t\to 0}{\sim} \frac{1}{t}P(z). \tag{1.17}$$

There are several PS methods, each of which based on different splitting kernels $\mathcal{K}$ and evolution variables t. In general, parton showers have leading-logarithmic (LL) and partial next-to-leading-logarithmic (NLL) accuracy, which means that they correctly take into account real emissions enhancement and virtual corrections at all orders of the kind $\alpha_S^n L^{2n}$ and $\alpha_S^n L^{2n-1}$ respectively, where $L \equiv \log(\varepsilon)$ is the logarithm associated to the IR limits. The LL term results from soft and collinear emissions, whereas NLL terms arise from hard collinear emissions and from the coherent treatment of soft emissions.

The natural following step is to consistently combine FO predictions with PS methods, since the former provide an accurate description of hard wide-angle emissions and non-logarithmic terms, while the latter describe the complementary soft-collinear regime.

Such combination is not a trivial task, and the naive procedure of simply applying a parton shower to a FO-generated event does not work in general. Contrary to parton showers, which give the (approximate) probability that any configuration with a certain well-defined number n of partons has been generated *exclusively* to all orders in α_S, FO matrix elements give the probability of producing *at least* n partons in a state calculated to a fixed α_S order, i.e. they are *inclusive* to any additional radiation. This also implies that, in case of combination with calculations that already include contributions of emissions from $n+1$ phase spaces (such as the real emission in NLO), we have to avoid to double count some regions of the phase space.

Various strategies have been devised to properly combine MEs and PS, which can be categorised into two groups.

- *matching* algorithms, where high-order correction to inclusive processes are consistently combined with a parton shower.

- *merging* algorithms, where parton shower and matrix elements with different jet multiplicities are combined in one inclusive sample in such a way that all hard jets (up to a certain multiplicity) that are emitted above a certain resolution scale (merging scale) are described in terms of matrix elements.

In the following we discuss two important examples of these categories: S-MC@NLO and MEPS.

1.3 Matching and merging methods

In this section we consider two methods that have been used throughout this work in order to combine perturbative calculations with parton showers. In the first part we discuss the matching of fixed-order matrix elements to parton showers and its extension to NLO accuracy, focussing on the S-MC@NLO method. In the second part we review the MEPS merging algorithm implemented in SHERPA.

1.3.1 The S-MC@NLO matching method

The first application of the PS method is to LO predictions. Following the notation used in the previous section, the differential LO+PS cross sections reads

$$d\sigma_{LO}^{PS} = d\Phi_B B(\Phi_B) \left[\Delta(t_{IR}, \mu_Q^2) + \int_{t_{IR}}^{\mu_Q^2} d\Phi_1 \frac{\alpha_S}{2\pi} \mathcal{K}(\Phi_1) \Delta(t_1, \mu_Q^2) \right], \quad (1.18)$$

where t_{IR} is the infrared cutoff and μ_Q is the *resummation scale*, that is, the highest scale at which the parton shower can emit. The first term in the square brackets gives the probability that no emission occurs between the scales t_{IR} and μ_Q, whereas the second term represents the probability to emit a parton at a scale t_1 in the range $t_{IR} < t_1 < \mu_Q^2$. The full brackets integrates to one yielding the correct LO normalisation, which reflects the unitarity of the formalism. However, the shape of differential observables which are sensitive to extra emissions can be affected. The above equation describes only the first PS emission at a scale t_1 , and the full shower history is generated via iterative emissions at the scales $\mu_Q > t_1 > t_2 > \cdots > t_{IR}$.

The POWHEG and MC@NLO matching methods

In case of NLO calculations the matching to a PS is not as trivial as in the LO case. In general we need to combine MEs predictions and shower evolution keeping NLO accuracy and avoiding the double counting that arises in case of extra emissions. In the POWHEG method [16,17], this is achieved by using the ratio of real-emission to Born matrix elements as emission kernel $\mathcal{K}$ in (1.16), that is

$$\Delta(t,t')_{POW} = \exp\left\{-\int_t^{t'} d\Phi_1 \frac{R(\Phi_B,\Phi_1)}{B(\Phi_B)}\right\}, \tag{1.19}$$

together with the replacement

$$B(\Phi_B) \to \tilde{B}(\Phi_B) = B(\Phi_B)+V(\Phi_B)+I(\Phi_B)+\int_1 d\Phi_1 \left[R(\Phi_B,\Phi_1) - S(\Phi_B,\Phi_1)\right]. \tag{1.20}$$

The resulting matching formula for the first emission in the original formulation of the POWHEG method reads

$$\sigma_{NLO}^{POW} = \int d\Phi_B \tilde{B}(\Phi_B) \left[\Delta(t_{IR},\mu_Q^2)_{POW} + \int_{t_{IR}}^{\mu_Q^2} d\Phi_1 \frac{R(\Phi_B,\Phi_1)}{B(\Phi_B)} \Delta(t_1,\mu_Q^2)_{POW}\right] \tag{1.21}$$

and provides accuracy up to first order in α_S for both the total cross section and differential distributions only provided that the upper integration limit in the real emission term is extended up to the hadronic centre of mass energy in order to fill the entire phase space, $\mu_Q^2 = s$. This entails to extend parton-shower resummation far beyond its range of validity, which corresponds to the soft and collinear regions.

In general, this can be avoided by splitting the real matrix elements R into an infrared-singular ("soft") and an infrared-regular ("hard") part,

$$R = D^{(A)} + H^{(A)}, \tag{1.22}$$

and by restricting the resummation to the former. In this case (1.8) becomes

$$\begin{aligned} d\sigma_{NLO} = {} & d\Phi_B \left[B(\Phi_B) + V(\Phi_B) + I(\Phi_B)\right] \\ & + d\Phi_R \left[D^{(A)}(\Phi_R) - S(\Phi_R)\right] + d\Phi_R H^{(A)}(\Phi_R), \end{aligned} \tag{1.23}$$

and the resulting matching formula reads

$$\begin{aligned} \sigma_{NLO+PS} = {} & \int d\Phi_B \bar{B}(\Phi_B) \left[\bar{\Delta}(t_{IR},\hat{s}) + \int_{t_{IR}}^{\hat{s}} d\Phi_1 \frac{D^{(A)}(\Phi_B,\Phi_1)}{B(\Phi_B)} \bar{\Delta}(t_1,\hat{s})\right] \\ & + \int d\Phi_R H^{(A)}(\Phi_R), \end{aligned} \tag{1.24}$$

where

$$\bar{B}(\Phi_B) = B(\Phi_B) + V(\Phi_B) + I(\Phi_B) + \int_1 d\Phi_1 \left[D^{(A)}(\Phi_B, \Phi_1) - S(\Phi_B, \Phi_1) \right], \tag{1.25}$$

and the Sudakov form factor

$$\bar{\Delta}(t, t') = \exp\left\{ - \int_t^{t'} d\Phi_1 \frac{D^{(A)}(\Phi_B, \Phi_1)}{B(\Phi_B)} \right\}. \tag{1.26}$$

The full procedure guarantees that only the singular parts are resummed, and avoids the exponentiation of the regular (hard) piece.

Within the POWHEG method the separation (1.22) is achieved through an appropriate damping function that provides a smooth splitting of the real-emission phase space into a soft and a hard part [54].

The MC@NLO matching formalism [15], which was introduced before the POWHEG method, can be regarded as an alternative realisation of the general matching formula (1.24)–(1.26). In the MC@NLO framework the "soft" part of the real-emission matrix element is identified with the first-order expansion of the parton-shower emission kernel via

$$D^{(A)}(\Phi_R) = B(\Phi_B)\mathcal{K}(\Phi_1). \tag{1.27}$$

However, this does not allow one to capture the exact $\mathcal{O}(\alpha_S)$ singular behaviour of the first emission. In fact, soft-divergences associated with subleading colour configurations are not taken into account by the leading logarithmic and leading colour PS kernel $\mathcal{K}(\Phi_1)$, yielding an infinite result for (1.25).

This issue was overcome in [55] for the case of heavy flavour production, by the introduction of an additional ad-hoc factor to (1.25) which tends to zero as the emissions become soft. This problem, instead, has been consistently solved in [18], where a significant improvement of the MC@NLO formalism has been presented. This method is known as S-MC@NLO and is outlined in the following.

The S-MC@NLO method

A great simplification arises in the MC@NLO matching approach (1.24)–(1.27) if the infrared subtraction terms are used as shower kernel, that is

$$D^{(A)}(\Phi_R) = S(\Phi_R) = B(\Phi_B)\mathcal{K}(\Phi_1). \tag{1.28}$$

This procedure has been proposed in [18] and implemented in the Sherpa [28] Monte Carlo framework. We will refer to this scheme as S-MC@NLO (Sherpa-MC@NLO) method. In Sherpa both the parton shower and the subtraction terms are based on the Catani-Seymour dipoles [53].

The expectation value of an arbitrary infrared safe observables $\mathcal{O}$ in the S-MC@NLO scheme reads

$$\begin{aligned}\langle\mathcal{O}\rangle = \int d\Phi_B \bar{B}(\Phi_B) \Bigg[& \Delta(t_{IR},\mu_Q^2)\mathcal{O}(\Phi_B) \\ & + \sum_{\tilde{ij},\tilde{k}} \int_{t_{IR}}^{\mu_Q^2} d\Phi_1^{ij,k} \frac{D_{ij,k}(\Phi_B,\Phi_1^{ij,k})}{B(\Phi_B)} \Delta(t_1,\mu_Q^2)\mathcal{O}(\Phi_R) \Bigg] \\ & + \int d\Phi_R \left[R(\Phi_R) - \sum_{ij,k} D_{ij,k}(\Phi_R)\Theta(\mu_Q^2 - t_1) \right] \mathcal{O}(\Phi_R),\end{aligned} \tag{1.29}$$

where

$$\begin{aligned}\bar{B}(\Phi_B) = B(\Phi_B) + V(\Phi_B) + \sum_{\tilde{ij},\tilde{k}} I_{\tilde{ij},\tilde{k}}(\Phi_B) \\ + \sum_{\tilde{ij},\tilde{k}} \int_1 d\Phi_1^{ij,k} \left[D_{ij,k}(\Phi_B,\Phi_1^{ij,k})\Theta(\mu_Q^2 - t_1) - S_{ij,k}(\Phi_B,\Phi_1^{ij,k}) \right],\end{aligned} \tag{1.30}$$

can be interpreted as a Born-level cross section times a local K-factor, $D_{ij,k}$ are the Catani-Seymour dipoles related to the splitting $\tilde{ij} \to i,j$, with spectator parton k,

$$I_{\tilde{ij},\tilde{k}}(\Phi_B) = \int_1 d\Phi_1^{ij,k} S_{ij,k}(\Phi_B,\Phi_1^{ij,k})$$

are the integrated dipoles (mapped to the Born phase space) and the $\tilde{ij},\tilde{k}$ (i,j,k) indices run over all the possible splittings. The overall Sudakov form factor $\Delta(t,t')$ is defined as

$$\begin{aligned}\Delta(t,t') &= \prod_{\tilde{ij},\tilde{k}} \Delta_{\tilde{ij},\tilde{k}}(t,t') \\ &= \prod_{\tilde{ij},\tilde{k}} \exp\left\{ - \sum_{i=q,g} \int_t^{t'} d\Phi_1^{ij,k} \frac{D_{ij,k}(\Phi_B,\Phi_1^{ij,k})}{B(\Phi_B)} \right\}\end{aligned} \tag{1.31}$$

and it differs from the ordinary parton-shower Sudakov as the Catani–Seymour dipole terms $D_{ij,k}$ include full colour and spin correlations. This makes the

non-logarithmic remainder (third line in (1.29)) not divergent [18], solving the issue related to sub-leading colour soft divergences pointed out in the previous section. The event is then showered with leading colour PS evolution, respecting the emission scales already present, that is, the shower does not emit above the first emission scale, if occurred. If necessary, truncated parton-shower [26] emissions are inserted to retain the logarithmic accuracy of the parton shower. We now show that (1.29)–(1.31) guarantee NLO accuracy and the correct matching of the first shower emission. To this end, we expand the equation at order α_S obtaining [3]

$$\begin{aligned}\langle\mathcal{O}\rangle|_{\alpha_S} = &\int d\Phi_B \left[B(\Phi_B) + V(\Phi_B) + I(\Phi_B)\right]\mathcal{O}(\Phi_B)\\ &+ \int d\Phi_B \sum_i \int d\Phi_1^i \left[D_i(\Phi_B,\Phi_1^i)\Theta(\mu_Q^2 - t_1) - S_i(\Phi_B,\Phi_1^i)\right]\mathcal{O}(\Phi_B)\\ &+ \int d\Phi_R \left[R(\Phi_R) - \sum_i D_i(\Phi_R)\Theta(\mu_Q^2 - t_1)\right]\mathcal{O}(\Phi_R) + \langle\hat{\mathcal{O}}\rangle_{\text{corr}}\\ &+ O(\alpha_S^2),\end{aligned} \tag{1.32}$$

where the correction term

$$\begin{aligned}\langle\hat{\mathcal{O}}\rangle_{\text{corr}} &= \int d\Phi_B B(\Phi_B)\Bigg[\Big(\Delta(t_{IR},\mu_Q^2) - 1\Big)\mathcal{O}(\Phi_B)\\ &\qquad + \sum_i \int_{t_{IR}}^{\mu_Q^2} d\Phi_1^i \frac{D_i(\Phi_R)}{B(\Phi_B)}\Delta(t_1,\mu_Q^2)\mathcal{O}(\Phi_R)\Bigg]\Bigg|_{\alpha_S}\\ &= \int d\Phi_R \sum_i D_i(\Phi_R)\Theta(\mu_Q^2 - t_1)\big[\mathcal{O}(\Phi_R) - \mathcal{O}(\Phi_B)\big] + O(\alpha_S^2),\end{aligned} \tag{1.33}$$

takes into account the mismatch originating from the definitions of the observable in the Born and real phase space. Noting that

$$\begin{aligned}&\int d\Phi_B I(\Phi_B) + \int_1 d\Phi_1^i \left[D_i(\Phi_B,\Phi_1^i)\Theta(\mu_Q^2 - t_1) - S_i(\Phi_B,\Phi_1^i)\right]\\ &= \int d\Phi_R D_i(\Phi_B,\Phi_1^i)\Theta(\mu_Q^2 - t_1)\end{aligned} \tag{1.34}$$

[3] We are now omitting dipole outgoing indices for sake of clarity.

and that

$$\begin{aligned}&\int d\Phi_R \left[R(\Phi_R) - \sum_i D_i(\Phi_R)\Theta(\mu_Q^2 - t_1)\right]\mathcal{O}(\Phi_R) + \langle\hat{\mathcal{O}}\rangle_{\text{corr}} \\ &= \int d\Phi_R R(\Phi_R)\mathcal{O}(\Phi_R) - \sum_i D_i(\Phi_R)\Theta(\mu_Q^2 - t_1)\mathcal{O}(\Phi_B)\end{aligned} \tag{1.35}$$

we find an equation equivalent to the NLO formula (1.10), with the replacement

$$S_i(\Phi_R) \to D_i(\Phi_R)\Theta(\mu_Q^2 - t_1),$$

which means that the observable is NLO accurate.

The only new ingredient w.r.t. a FO calculation in this framework is the introduction of the resummation scale μ_Q. This scale controls the transition between the hard-ME region to the soft-PS one and sets the maximum scale at which the shower can emit. It is easy to check that in the limit $\mu_Q \to 0$ we formally recover the plain NLO fixed order cross section of (1.8).

In Section 2.1 we report a phenomenological application of this method to $t\bar{t}b\bar{b}$ production at the LHC.

Theoretical uncertainties

The expression (1.29) is NLO accurate in both total cross section and differential observables. Theoretical uncertainties are typically assessed through variations of various technical parameters that provide an indication of the size of higher-order contributions beyond NLO.

The factorisation scale μ_F and the renormalisation scale μ_R dependence is embodied in the $\bar{B}$ part and the hard remainder, and their variation is formally $O(\alpha_S^2)$ relative to the Born contribution. Renormalisation scale variations should be consistently applied also at the level of the first matched emission at both Sudakov level and shower radiation level. However, this is not possible in SHERPA yet[4], and thus in our simulations throughout this work we have kept these scales fixed and varied only the renormalisation scale in the $\bar{B}$ and $R - D$ terms in (1.29). The scales μ_R and μ_F are usually chosen close to the characteristic scale Q of the process at hand, and theoretical uncertainties are typically assessed by a factor two variation.

The resummation scale μ_Q sets the upper bound for the shower first emission. The dependence on this scale can be used to assess the uncertainties associated with the resummation procedure, which are typically quantified by a factor two (or $\sqrt{2}$) variation.

[4]This feature will be available in the next SHERPA public release.

The assessment of the uncertainties is crucial in phenomenological calculations. In this respect, the multi-scale nature of multi-particle processes poses nontrivial issues, since the choice of the scales is not unique. In these circumstances the typical up/down scale variation of the scales μ_R, μ_F, μ_Q might be not sufficient to correctly evaluate the uncertainties, but rather different dynamic scales should be compared to better account for uncertainties related to the multiscale nature of the problem. In section 2.2.2 we report a study on the theoretical uncertainties related to the $t\bar{t}b\bar{b}$ process at LHC using the S-MC@NLO formalism.

1.3.2 Multi-jet merging

In this section we outline the key aspects of another kind of procedure for the combination of MEs and PS, namely the merging algorithms. As already mentioned above, these strategies are based on the combination of final states with different jet multiplicities in one inclusive sample, in such a way that all hard jets (up to a certain multiplicity) that are emitted above a certain resolution scale (merging scale) are described in terms of matrix elements.

There are several merging methods on the market, such as CKKW [21,22], MLM [23], FxFx [24], and MEPS [19,20,26,56,57]. In the following, we focus on a particular implementation of such algorithms, namely MEPS@LO for LO merging and its NLO extension, MEPS@NLO, both implemented in SHERPA.

Multi-jet merging at LO: MEPS@LO

The main idea behind merging algorithms is to combine into a single sample matrix elements with different jet multiplicities up to a certain maximum number of jets N_{max}, each matched to the parton shower, in such a way that both the (N)LO and the shower logarithmic accuracies are preserved. To this end, the phase space is separated in a hard region, filled by MEs, and a soft region, filled by the PS. The separation is implemented through a merging scale defined in terms of a k_T-type measure, called Q_{cut}. In this manner each multi-jet event is described, in the hard region, by matrix elements of appropriate jet multiplicity, whereas extra shower emissions give rise to the jet substructure without leading to additional jets above the merging scale.

The MEPS@LO algorithm [26, 56, 57] is based on the interpretation of a LO ME for $pp \to X + k$–partons in terms of the most likely branching history, that is, the most likely sequence of (ordered) emissions from which such event might have been generated, starting from a core process $pp \to X$ that occurs at the core scale μ_{core}. The branching history, and the related nodal scales $t_1 < \cdots < t_n < \mu_{\text{core}}$ associated to the event at hand, are determined

by inverting the SHERPA parton shower. More precisely, the actual shower history is determined in a probabilistic way based on the relative probabilities computed with the SHERPA parton shower for all possible histories that connect a given multi-parton event with the core process. The shower logarithmic accuracy is then preserved above the merging scale $Q_{\rm cut}$ through a CKKW-like scale choice [21, 22] and by dressing MEs with Sudakov form factors (corresponding to the SHERPA parton shower) based on the actual branching history.

Let us now consider the case of the merging at LO of two multiplicities, n and $n+1$. In each phase space the separation of the hard region from the soft region is implemented in terms of a merging cut, $Q_{\rm cut}$, in the k_T-jet measure. More precisely, each k-parton final state is characterised by ordered branching scales $Q_1 > Q_2 > \cdots > Q_k$, which are computed via k_T-jet clustering and are used in order to determine the number of resolved emissions according to the $Q_{\rm cut}$ scale. For instance, if $Q_n > Q_{\rm cut} > Q_{n+1}$ the first n emissions are handled as resolved and the remaining ones as unresolved. In general $Q_{\rm cut}$ should be chosen smaller than the minimum jet transverse momenta in order to guarantee that each potentially observed jet is treated as resolved and thus described in terms of matrix elements.

The MEPS@LO master equation in this case reads

$$\begin{aligned}\langle\mathcal{O}\rangle_{\text{MEPS@LO}} = \int d\Phi_n B_n \Bigg[& \Delta_n(t_{IR},\mu_Q^2)\mathcal{O}_n \\ & + \int_{t_{IR}}^{\mu_Q^2} d\Phi_1 \mathcal{K}_n \Delta_n(t_{n+1},\mu_Q^2)\Theta(Q_{\rm cut}-Q_{n+1})\mathcal{O}_{n+1}\Bigg] \\ + \int d\Phi_{n+1} B_{n+1} & \Delta_n(t_{n+1},\mu_Q^2)\Theta(Q_{n+1}-Q_{\rm cut}) \\ \times \Bigg[& \Delta_{n+1}(t_{IR},t_{n+1})\mathcal{O}_{n+1} + \int_{t_{IR}}^{t_{n+1}} d\Phi_1 \mathcal{K}_{n+1}\Delta_{n+1}(t_{n+2},t_{n+1})\mathcal{O}_{n+2}\Bigg],\end{aligned} \tag{1.36}$$

where B_k, $\mathcal{O}_k$, $\mathcal{K}_k$ and Δ_k are, respectively, the Born MEs, the observable, the parton shower kernel and the related Sudakov form factor defined in the k-particle phase space. As mentioned above, the Born MEs are supplemented by a CKKW-like scale choice, implemented by evaluating the matrix elements at $\mu_R = \mu_{\rm core}$ and applying a subsequent reweighing factor $\prod_l \alpha_S(t_l)/\alpha_S(\mu_R^2)$, where l runs over all the nodal scales. In comparison with the LO+PS equation (1.18), the shower is restricted to emit below $Q_{\rm cut}$ due to the step function Θ in the n-jet process, whereas the hard matrix element in the $n+1$ phase space B_{n+1} fills the phase space above $Q_{\rm cut}$ and gets a Sudakov form

factor that matches the one of the parton shower emission, thereby preserving the logarithmic accuracy of parton-shower resummation in the region above $Q_{\rm cut}$ and ensuring a mild $Q_{\rm cut}$ dependence of the predictions. In fact, in the limit of small merging scale, the logarithmic $Q_{\rm cut}$ dependence of MEPS@LO predictions is beyond the logarithmic accuracy of the parton shower.

MEPS@NLO

This multi-jet merging algorithm has been extended to NLO (MEPS@NLO) in [19, 20]. In this case S-MC@NLO simulations of different jet multiplicities up to a maximum number N_{max} are combined together in an inclusive sample, and a merging cut $Q_{\rm cut}$ is used to separate the hard region from the soft region for each emission. In analogy to the LO case, an n-particle NLO+PS contribution of type (1.29) fills the region below the merging cut $Q_{n+1} < Q_{\rm cut}$ according to,

$$\begin{aligned}\langle \mathcal{O}_n \rangle_{\rm NLO+PS} = \int d\Phi_n \bar{B}_n \Bigg[& \Delta_n(t_{IR}, \mu_Q^2)\mathcal{O}_n \\ & + \int_{t_{IR}}^{\mu_Q^2} d\Phi_1 \frac{D_n}{B_n} \Delta_n(t_{n+1}, \mu_Q^2)\Theta(Q_{\rm cut} - Q_{n+1})\mathcal{O}_{n+1} \Bigg] \\ + \int d\Phi_{n+1} \left[R_n - D_n\right] & \Theta(Q_{\rm cut} - Q_{n+1})\mathcal{O}_{n+1},\end{aligned} \tag{1.37}$$

while the region above the merging cut, $Q_{n+1} > Q_{\rm cut}$, is filled by an S-MC@NLO simulation with one additional jet. The latter is supplemented by a Sudakov form factor $\Delta_n(t_{n+1}, \mu_Q^2)$, which renders it exclusive with respect to additional emissions harder than t_{n+1}, thereby avoiding double counting with contributions from the phase space with 2 additional partons in the $n+2$ phase space. Besides that, at NLO one has to avoid the double counting between the Sudakov form factor and related contributions that enter the $\bar{B}$ terms through the virtual corrections. This is achieved by the addition of the shower counterterms of the form

$$\bar{B}_{k+1}\, \Delta_k(t_{k+1}, \mu_Q^2) \longrightarrow \bar{B}_{k+1} \left[1 + \frac{B_{k+1}}{\bar{B}_{k+1}} \int_{t_{k+1}}^{\mu_Q^2} d\Phi_1 \mathcal{K}_{k+1}\right] \Delta_k(t_{k+1}, \mu_Q^2), \tag{1.38}$$

which cancels the first order expansion contribution of the Sudakov form factor. Finally, the MEPS@NLO master formula for the merging of n and

$n+1$ particle contributions reads

$$
\begin{aligned}
\langle\mathcal{O}\rangle_{\text{MEPS@NLO}} = \int d\Phi_n \bar{B}_n \Bigg[& \Delta_n(t_{IR},\mu_Q^2)\mathcal{O}_n \\
& + \int_{t_{IR}}^{\mu_Q^2} d\Phi_1 \frac{D_n}{B_n} \Delta_n(t_{n+1},\mu_Q^2)\Theta(Q_{\text{cut}}-Q_{n+1})\mathcal{O}_{n+1} \Bigg] \\
+ \int d\Phi_{n+1} & \left[R_n - D_n\right] \Theta(Q_{\text{cut}}-Q_{n+1})\mathcal{O}_{n+1} \\
+ \int d\Phi_{n+1} \bar{B}_{n+1} & \left[1 + \frac{B_{n+1}}{\bar{B}_{n+1}} \int_{t_{n+1}}^{\mu_Q^2} d\Phi_1 \mathcal{K}_n \right] \Delta_n(t_{n+1},\mu_Q^2)\Theta(Q_{n+1}-Q_{\text{cut}}) \\
& \times \left[\Delta_{n+1}(t_{IR},t_{n+1})\mathcal{O}_{n+1} + \int_{t_{IR}}^{t_{n+1}} d\Phi_1 \frac{D_{n+1}}{B_{n+1}} \Delta_{n+1}(t_{n+2},t_{n+1})\mathcal{O}_{n+2} \right] \\
+ \int d\Phi_{n+2} & \left[R_{n+1} - D_{n+1}\right] \Delta_n(t_{n+1},\mu_Q^2)\Theta(Q_{n+1}-Q_{\text{cut}})\mathcal{O}_{n+2},
\end{aligned}
\tag{1.39}
$$

where, analogously to the LO case, all the MEs are evaluated at scale $\mu_R = \mu_{\text{core}}$ and reweighed by factors $\prod_l \alpha_S(t_l)/\alpha_S(\mu_R^2)$. Note that in (1.39) the S-MC@NLO simulation with $n+1$ partons is treated inclusively with respect to extra emissions, i.e. no merging cut is applied to Q_{n+2}. In general, the MEPS@NLO procedure can be extended to higher jet multiplicities by applying a cut $Q_{n+2} < Q_{\text{cut}}$ to the simulation with $(n+1)$ partons and filling the region $Q_{n+2} > Q_{\text{cut}}$ with an extra S-MC@NLO simulation with $n+2$ partons supplemented with appropriate Sudakov form factors and shower counterterms, and so on.

1.4 The MINLO method

In this section we introduce the MINLO (Multi-scale improved NLO) procedure and present its fully automated implementation in Sherpa. The MINLO method [38] provides NLO accuracy as well as resummation of soft and collinear logarithms that arise in the presence of large ratios of scales. The starting point of the MINLO approach, as in merging algorithms, is the interpretation of a multi-jet event $pp \to X + n$ jets as a core process plus a sequence of n jet emissions, according to the most likely branching history. Once such branching history has been determined, MEs are supplemented by a CKKW-like [21, 22] renormalisation scale and appropriate Sudakov form factors, which render the n emissions exclusive with respect to any harder

emission. Thus, from the viewpoint of Sudakov resummation, MINLO events with n emissions can be regarded as the outcome of a parton shower after a fixed number n of branchings.

This section is organized as follows: in the first part we present a LO implementation of the MiNLO algorithm (that we call "MiLO"), introducing the notation and focusing on the differences of our implementation with respect to the original proposal [38]. In the second part we outline the NLO extension of such algorithm and finally we present a validation study, where we compare our MINLO implementation to the Meps procedure.

The algorithm outlined in this section has been encoded in Sherpa in a fully general way, which renders it applicable to any SM process of arbitrary jet multiplicity, and will be publicly available in the near future. A phenomenological application of this code to $t\bar{t}$+multijet production with up to three jets at NLO accuracy has been recently published in [39] and is reported in Chapter 3.

1.4.1 Implementation of MiNLO at LO

Here we introduce our LO implementation of the MiNLO algorithm, which corresponds to CKKW approach [21, 22] applied to a fixed-multiplicity calculation.

Let us consider a generic process $pp \to X+N$ partons, where X is the core configuration that cannot be further clustered. Such process is characterized by a α_S power $P = M + N$ at LO, where M is the intrinsic QCD order of the core process $pp \to X$ (for instance, $M = 0$ for vector boson production, $M = 2$ for $gg \to H$ and $t\bar{t}$ production). The MiLO algorithm consists of the following steps.

i) By means of the k_T jet algorithm [58], we recursively cluster each $X+N$-parton event back to a $pp \to X$ core process, and we determine the k_T^2 scales $t_1 < \cdots < t_N$ of the related branchings. For each Final-Final (FF) branching, $\widetilde{ij} \to ij$, we take the scale

$$t = 2\min(k_{iT}^2, k_{jT}^2)\frac{\cosh(Y_i - Y_j) - \cos(\phi_i - \phi_j)}{R^2}, \tag{1.40}$$

while for each Final-Initial (FI) branching we use

$$t = k_{iT}^2, \tag{1.41}$$

where k_{iT}, Y_i and ϕ_i are the transverse momentum, the rapidity and the azimuthal angle respectively of the parton i, and R is a freely adjustable parameter. Only clusterings which are compatible in flavours

are allowed, that is, gluons with gluons, giving gluon pseudo-partons, gluons with quarks, giving pseudo-quarks with the same flavour, and quarks with antiquarks with same flavour, giving pseudo-gluons [5]. To each clustered pseudoparton we assign a label according to how it has been generated, i.e. we call it Final (F) if it originates from two final pseudo-partons and Initial (I) otherwise. The purpose of this label will be explained later.

We call *skeleton* of the event the sequence of clusterings yielded by the algorithm. An example of skeleton of an event is illustrated in Fig. 1.1. To each skeleton we assign a core scale, which corresponds to the characteristic energy scale of the core process and is defined, in general, in terms of the corresponding clustered momenta. There are situations where the core scale μ_{core} is not the hardest scale of the event. In particular, this happens in case of jet emissions harder than μ_{core}. As only ordered emissions can be resummed, for each event we only consider ordered emissions and we handle unordered ones as part of the core process. Thus we always deal with ordered skeletons, with

$$t_1 < \cdots < t_n < \mu^2_{\text{core}}, \tag{1.42}$$

where, in general, $n \leq N$ and the core configuration can contain one or more unclustered pseudopartons, thus having an intrinsic QCD order $m = M + N - n \geq M$. More details on the treatment of unordered emissions are provided in Subsection 1.4.3.

ii) The $m \geq M$ powers of α_S associated with the core process are evaluated at the scale μ_{core}, while the remaining n powers, associated to the extra jet emissions, are evaluated at the related branching scales $\mu_i = \sqrt{t_i}$. This choice is equivalent to setting the effective renormalisation scale μ_{eff}, defined by

$$\left[\alpha_S(\mu^2_{eff})\right]^{m+n} = \left[\alpha_S(\mu^2_{\text{core}})\right]^m \prod_{i=1}^{n} \alpha_S(t_i). \tag{1.43}$$

Renormalisation scale uncertainties are assessed by setting $\mu_R = \xi_R \mu_{eff}$, where ξ_R is the renormalisation scale factor, which is typically varied between 0.5 and 2.

iii) We identify the factorisation scale with the branching scale of the softest emission by setting

$$\mu_F = \xi_F \sqrt{t_1}, \tag{1.44}$$

[5] The particular case of top recombination will be discussed in Chapter 3.

where ξ_F is applied for factorisation scale variations. The choice (1.44) is motivated by the fact that, as discussed below, matrix elements are supplemented by Sudakov form factors that effectively resum initial-state collinear logarithms down to a resolution scale t_1. This scale should thus be regarded as a matching scale between logarithmic resummation via analytic Sudakov form factors (FFs) and PDFs.

iv) We dress matrix elements with Sudakov form factors associated with the internal and external lines of the event skeleton.

Explicit analytic formulas for the employed form factors and additional details can be found in Appendix A. The employed splitting functions account for quark mass effects [59–61] and include two-loop cusp terms [51].

For each external line connected to a nodal scale t_j we apply a Sudakov form factor

$$\Delta_{f_j}^{ext}(t_j;\mu_{res}^2) \equiv \frac{\Delta_{f_j}(\mu_{res}^2,t_j)}{\Delta_{f_j}(\mu_{res}^2,t_1)} \tag{1.45}$$

where f_j is its flavour ($f_j = g,q,\bar{q}$), t_j is the nodal scale the line ends to, t_1 is the lowest recombination scale, and μ_{res} is the resolution scale defined as

$$\mu_{res}^2 = \begin{cases} t_1 & \text{if the line is final (F)} \\ \mu_F^2 = \xi_F^2 t_1 & \text{if the line is initial (I)} \end{cases} \tag{1.46}$$

This is slightly different than the original proposal, where the denominator was set to one in (1.45) and there was no distinction between initial and final external lines. In our case $\Delta_{f_l}(\mu_{res}^2,t_1) = \Delta_{f_l}(t_1,t_1) = 1$ only for $\xi_F = 1$. Such distinction has been introduced in order to match the logarithmic resummation occurring in the PDFs with the one provided by the Sudakov FFs in case of factorisation scale variations. In Fig. 1.2 a pictorial representation of the behaviour of the algorithm in case of a μ_F variation is shown. Sudakov form factors (1.45) can be interpreted as the probability for the flavour f_j to not emit any parton between the scales t_j and t_1.

For each internal line joining two nodal scales $t_i > t_j$ we apply a Sudakov form factor equal to

$$\Delta_{f_{ij}}^{int}(t_j,t_i;\mu_{res}^2) \equiv \frac{\Delta_{f_{ij}}(\mu_{res}^2,t_i)}{\Delta_{f_{ij}}(\mu_{res}^2,t_j)} \tag{1.47}$$

where f_{ij} is the flavour of such internal line. The interpretation of this part is straightforward: each line of flavour f connecting two scales,

t_i and t_j, is associated to a no-emission probability between those two scales. Again, we make the distinction between initial and final state lines using the definition (1.46) of μ_{res}.

The dressing with appropriate Sudakov FFs of the internal and external lines renders the n emissions similarly exclusive as if they were produced by a shower stopped after n emissions, i.e. exclusive with respect to any emission harder than the n^{th} one.

What differs in the MiLO approach from the standard CKKW method is the choice of the resolution scale of the process. While the CKKW approach implements a fixed resolution (merging) scale, in MiLO such scale is chosen on event-per-event basis and corresponds to the lowest clustering scale of the skeleton. Each event should be regarded as unresolved below its lowest scale μ_{res}. Thus, at LO, it is possible to obtain a fully exclusive simulation by simply showering MiLO events using and event-by-event shower with starting scale $\mu_Q = \sqrt{t_1}$[6] . However, at NLO the matching of MiNLO events to the parton shower is non trivial. Thus in this thesis MiLO and MiNLO predictions will not be matched to parton showers.

The MiLO approach can be summarized in the following formula

$$|M_{\text{MiLO}}|^2 = |M_{LO}(\mu_F^2, \mu_{eff}^2)|^2 \times \prod_l \Delta_{f_l}^{ext}(t_l; \mu_{res}^2) \prod_{ij} \Delta_{f_{ij}}^{int}(t_j, t_i; \mu_{res}^2), \quad (1.48)$$

where $M_{LO}(\mu_{eff}^2)$ is the LO fixed-order matrix element calculated with renormalisation scale $\mu_R = \mu_{eff}$ and factorisation scale μ_F, the index l runs over all external lines of flavour f_l connected to node t_l, and the second product runs over all internal lines of flavour f_{ij} connecting two scale $t_i > t_j$.

1.4.2 Extension to NLO: the MiNLO algorithm

Two issues must be addressed moving to NLO. The first one is to implement the scale-setting procedure and Sudakov FFs in such a way that preserves the pole cancellations that take place between the Born and the real phase space. Second, we must remove the $\mathcal{O}(\alpha_S)$ contribution generated by the Sudakov form factors in order to avoid double counting of NLO effects. In the following we will detail the modifications required to extend the MiLO algorithm to NLO.

Let us consider a fixed-order NLO event in the subtraction formalism, as defined in Eq. (1.8). Such event is characterised by a Born and a virtual

[6] In principle, one should also take care of the mismatch between the different evolution variables in the Sudakov FFs and the parton shower by means of a truncated shower.

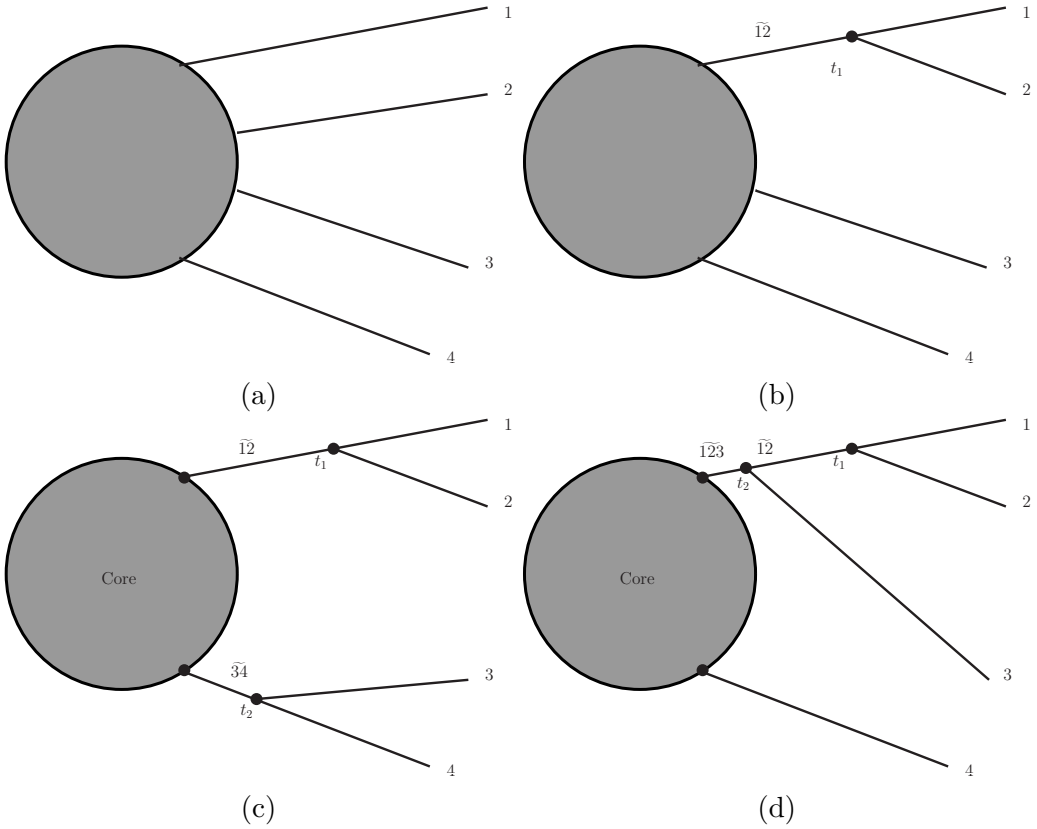


Figure 1.1: In figure (a) a generic $2 \to 4$ partonic process is shown, and in (b) how it looks after one clustering. Figures (c) and (d) represent two possible branching histories of the event.

weight, living in the Born phase space (with N partons), a real-emission term, defined in the real phase space (with $N+1$ partons), as well as real-emission subtraction terms, that live in both phase spaces.

i) As in step i) of the MiLO algorithm we apply the k_T algorithm in order to determine the skeleton of the event. The contributions that live in the Born phase space, i.e. the Born (B) and one-loop contributions (V), as well as all IR-subtraction terms (I and S), are handled exactly as in LO, whereas in case of a real contribution (R), we drop the first clustering node (that we will call t_0), and consider only the underlying Born event. This clustering corresponds to the softest (unresolved) real radiation and must be ignored in the scale setting procedure and in the Sudakov FFs in order to guarantee a consistent cancellation of the infrared poles.

ii)-iv) Same as the MiLO method steps ii)-iv).

v) In case of Born events, we subtract the expansion of the Sudakov form factors at NLO with the replacement

$$B \Rightarrow B_{\text{corr}} \equiv B \times \left(1 - \sum_{ij} \left[\Delta^{(1)}_{f_{ij}}(\mu^2_{res}, t_i) - \Delta^{(1)}_{f_{ij}}(\mu^2_{res}, t_j)\right] - \sum_{l} \left[\Delta^{(1)}_{f_l}(\mu^2_{res}, t_l) - \Delta^{(1)}_{f_l}(\mu^2_{res}, t_1)\right]\right), \tag{1.49}$$

where the index ij runs over all the pair of nodes (t_i, t_j) connected by a line of flavour f_{ij}, the index l runs over all the external lines connected to their node t_l and $\Delta^{(1)}$ is the $\mathcal{O}(\alpha_S)$ expansion of the Sudakov form factor. We emphasise that, in case of factorisation scale variation, the resolution scale μ_{res} in each correction term $\Delta^{(1)}$ is set consistently to the resummed Sudakov form factor Δ .

vi) The $(n+m+1)^{th}$ power of the coupling constant present in the real and virtual cross section, in all the IR subtraction terms, as well as in the Born subtraction term (1.49), is set to

$$[\alpha_S]^{n+m+1} = \left\{ \left[\alpha_S(\mu^2_{\text{core}})\right]^m \prod_{i=1}^{n} \alpha_S(t_i) \right\}^{\frac{n+m+1}{n+m}}, \tag{1.50}$$

where the n clustering scales are determined in the Born or real-emission phase space, depending on the contribution. The choice of such scale differs from the original implementation, in which the $(n+m+1)^{th}$ power of the α_S value was set to the arithmetic mean of the other coupling constants. However the effect of a different scale is formally a NNLO effect, and thus beyond the accuracy of the algorithm.

The procedure outlined above is generic and thus it can be implemented in any MC generator and applied to a wide class of processes.

1.4.3 Technical aspects of the implementation

In the following we discuss some additional aspects that differ from the original proposal of [38].

Ordered scales

In general the k_T jet algorithm is known to yield ordered scales, as it is actually based on the ordering of the k_T measure (1.40)–(1.41). However

there are two aspects in the MiNLO procedure that can spoil the ordering of the skeleton. The first aspect is the choice of the core scale $\mu_{\rm core}$. There are cases in which $\mu_{\rm core}$ is not the hardest scale in the event. For instance, this can happen in vector-boson+jets production, where an additional jet is emitted at high p_T, at a scale that is harder than the typical core process scale of order $\mu_{\rm core} \sim M_V$. In the original MiNLO proposal this issue is fixed by setting $\mu^2_{\rm core} = t_n$, where t_n is the hardest scale of the event. However this choice results suboptimal in processes where the core process itself is QCD induced, such as $t\bar{t}$+multijet production. In this case the core process is $\mathcal{O}(\alpha_S^2)$ and the emission of a very hard jet would imply that three powers of α_S are evaluated at such a hard scale. The second aspect that can spoil the ordering in the MiNLO algorithm is the presence of forbidden clusterings. As the procedure is not flavour blind, there are configurations that lead to softer scales in advanced steps of the k_T algorithm. An example is the (sub) process $q'\,\bar{q}' \to X + q\,\bar{q}$, with $q \neq q'$. The first clustering at scale t_1 is constraint to be $g \to q\,\bar{q}$, leading to the configuration $q'\,\bar{q}' \to X + g$. Here the gluon may cluster to either beams, but the new scale t_2 is not guaranteed to be higher than t_1. Therefore we adopted a specific prescription to enforce ordered skeletons. In the first place we required that a clustering can occur only if

$$t_{i+1} > t_i.$$

This avoids the unordering due to the flavour dependence. As for the unordering due to the core scale, whenever the condition

$$\mu^2_{\rm core} > t_N \tag{1.51}$$

is not satisfied, the last clustering is unfolded, an alternative N-th clustering is attempted, and the condition (1.51) is tested again. If the condition is satisfied, the new clustered configuration is accepted, otherwise a different clustering is considered, if available. If no other clustering can be attempted, the unfolded configuration is then considered as core process, and the condition $\mu^2_{\rm core} > t_{N-1}$ is tested. This prescription is then iterated until an ordered skeleton with $t_1 < \cdots < t_n < \mu^2_{\rm core}$ is obtained.

Initial state clustering

There are events where a final pseudoparton is allowed by flavour configuration to cluster with either beams. In such cases the k_T of the parton (1.41) is an ambiguous measure, as it does not discriminate any preferred clustering. In these situations we perform the clustering which minimises the invariant mass of the system.

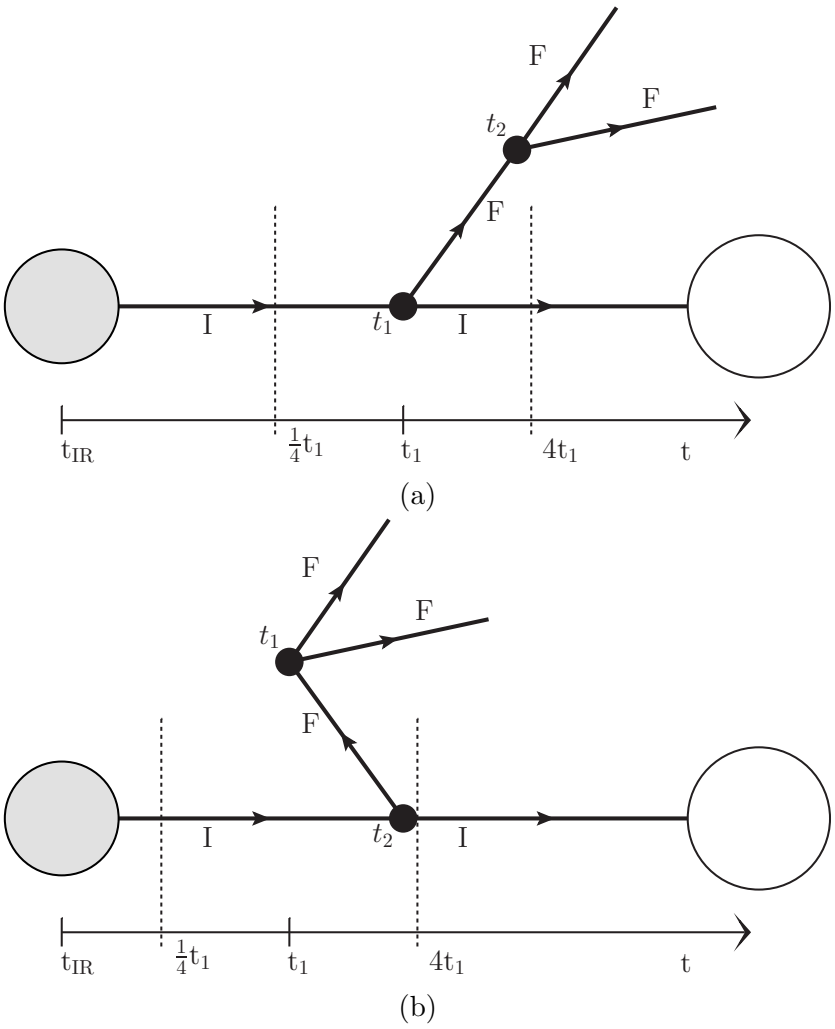


Figure 1.2: Representation of the interplay between μ_{res} and μ_F in case of FI and FF clusterings. The left blob is an incoming hadron, the right blob is the core process, and each initial(final) line is labelled as I(F). Dashed lines represent the effect of μ_F variations in the t axis. In (a) a FI clustering at scale t_1 is followed by a FF one at scale t_2, while in (b) a FF clustering precedes a FI clustering. In both examples, the variation $\mu_{res} = \mu_F = \sqrt{t_1}/2$ is necessary to match the evolution of PDFs for IS splittings from the cutoff t_{IR} to the core scale. In the case of $\mu_{res} = \mu_F = 2\sqrt{t_1}$ all the inverted emissions associated to FI clusterings are not resummed if $t_i < \mu_F^2$, as they are already present in the PDFs.

1.4.4 Validation against MEPS merging

The MiNLO algorithm described above has been implemented in SHERPA in a fully generic way, so that it can be applied to any process. Sudakov form factors have been coded with the formulas in Appendix A, taking into account also heavy-quark mass effects [59–61], power corrections in the evolution variable to LO DGLAP equations and the 2-loops α_S cusp term [51]. We have chosen $R = 0.57$ in the k_T algorithm, but this parameter is freely adjustable. The MiNLO algorithm has been encoded in such a way that Sudakov form factors are calculated and applied on the fly during event generation. Thanks to the resulting Sudakov damping of infrared singularities we have been able to produce events efficiently with very low IR cuts (order of 1 GeV), allowing to test our prescription down to very low scales with good statistic. This has allowed us to test the consistent resummation of logarithmic effects in the Sudakov peak region for several processes.

We tested our algorithm with $W + n$ jets processes, with off-shell W, focusing on the systematic comparison between our MI(N)LO implementation and MEPS@(N)LO method implemented in SHERPA. The goal of the comparison is to test the consistency of Sudakov resummation, therefore we performed our analysis at parton level, i.e. without hadronisation, UE and hadron decays, focusing on the behaviour of jet emissions in the soft regions. In general, at high transverse momenta, we expect significant deviations between MEPS and MiNLO due to various differences in the two methods. The first difference is that, contrary to MEPS@(N)LO, MI(N)LO predictions are not showered. This implies that, in the latter, the effect of multiple emission is only implemented in the form of Sudakov form factors that render MI(N)LO events with fixed parton multiplicity exclusive with respect to additional QCD emissions. More precisely, each event is characterised by a resolution scale that corresponds to the k_T of its softest parton, below which the jet substructure is not explicitly resolved. In MEPS, instead, such substructure is generated by the parton shower, that is allowed to fill the phase space with emissions down to the hadronisation scale. While such effects are suppressed in regions where the jet transverse momenta are small, for hard events they can become very important, and they can bias the comparison in a significant way. The second main difference is related to the clustering procedure of the two algorithms. Depending on the kinematic region, the MEPS approach can cluster Vjj topologies back to either Drell-Yan configurations (preferred in regions with soft jets), Vj configurations (in the region of intermediate jet p_T) or dijet configurations (in the regions with hard jets), whereas MiNLO can cluster only QCD partons back to a $V+$ jets configuration. The resulting differences in the core scale settings can give rise to

sizeable discrepancies between the two methods in the description of hard jet emissions. For these reasons in this comparison we will focus on the soft regions, where the comparison of MEPS and MINLO predictions will allow us to verify the consistent treatment of Sudakov logarithms in our MINLO implementation.

Setup of the test

We study the processes $pp \to e^- \bar{\nu}_e$+jets at 13 TeV. A similar process has been investigated with the MINLO algorithm in [62], including also comparisons with experimental data. We show MEPS@LO, MEPS@NLO, MILO and MINLO predictions for selected jet multiplicities. In particular, we performed the following simulations: MEPS@(N)LO $W + n \leq 2$ jets (which we denote $Wjj_{\texttt{MEPS@(N)LO}}$), Wj and Wjj using MI(N)LO methods (called $Wj_{\texttt{MI(N)LO}}$ and $Wjj_{\texttt{MI(N)LO}}$ respectively), for a total of 6 predictions. Events have been generated with SHERPA, whereas we used OPENLOOPS for NLO virtual amplitudes. Analyses are carried out with the help of RIVET [63]. The W^- boson has a mass of $M_W = 80.385$ GeV and a total width of $\Gamma_W = 2.085$ GeV, for the Z boson we set a mass of $M_Z = 91.1876$ GeV. We use in all our simulations the CT14 NLO [64] parton distribution functions. MI(N)LO simulations require a technical cutoff on the MEs in the IR limits. To this end, using the k_T algorithm with $R = 0.4$, we required the presence of at least 1(2) jets with $p_T > 2$ GeV for the $Wj(j)$ simulation. This avoids the region around the Landau pole, but at the same time could introduce some biases in the extremely soft region. In order to make a fair comparison of the algorithms in the Sudakov peak region we set all the internal shower parameters in MEPS affecting the evaluation of α_S consistently. Specifically, we set all the rescaling factors y of the kind $\alpha_S(y\,t)$ to 1, and in the MILO and MEPS@LO simulations the option of resumming subleading logarithms of Catani-Marchesini-Webber kind [51] through an appropriate rescaling of the α_S argument was deactivated. Regarding the scale choices, both algorithms rely on a CKKW-like scale, which must be supported by a core process scale, that we set for both to be

$$\mu_{\text{core}} = \frac{1}{2} H_T, \qquad H_T \equiv \sum_{i=W,jets} E_{T,i} \tag{1.52}$$

where $E_{T,i} = \sqrt{M_i^2 + p_{T,i}^2}$ is the transverse mass of the object i, and the sum includes only those (unordered) jets that are included in the core process. Moreover, we needed to specify the merging scale for the MEPS algorithm, which we set to $Q_{\text{cut}} = 20$ GeV. At analysis level, we consider jets obtained

with an anti-k_T algorithm [65] and jet radius $R = 0.4$ with a pseudorapidity cut of $|\eta_j| \leq 2.5$. We do not impose any p_T cut on the jets.

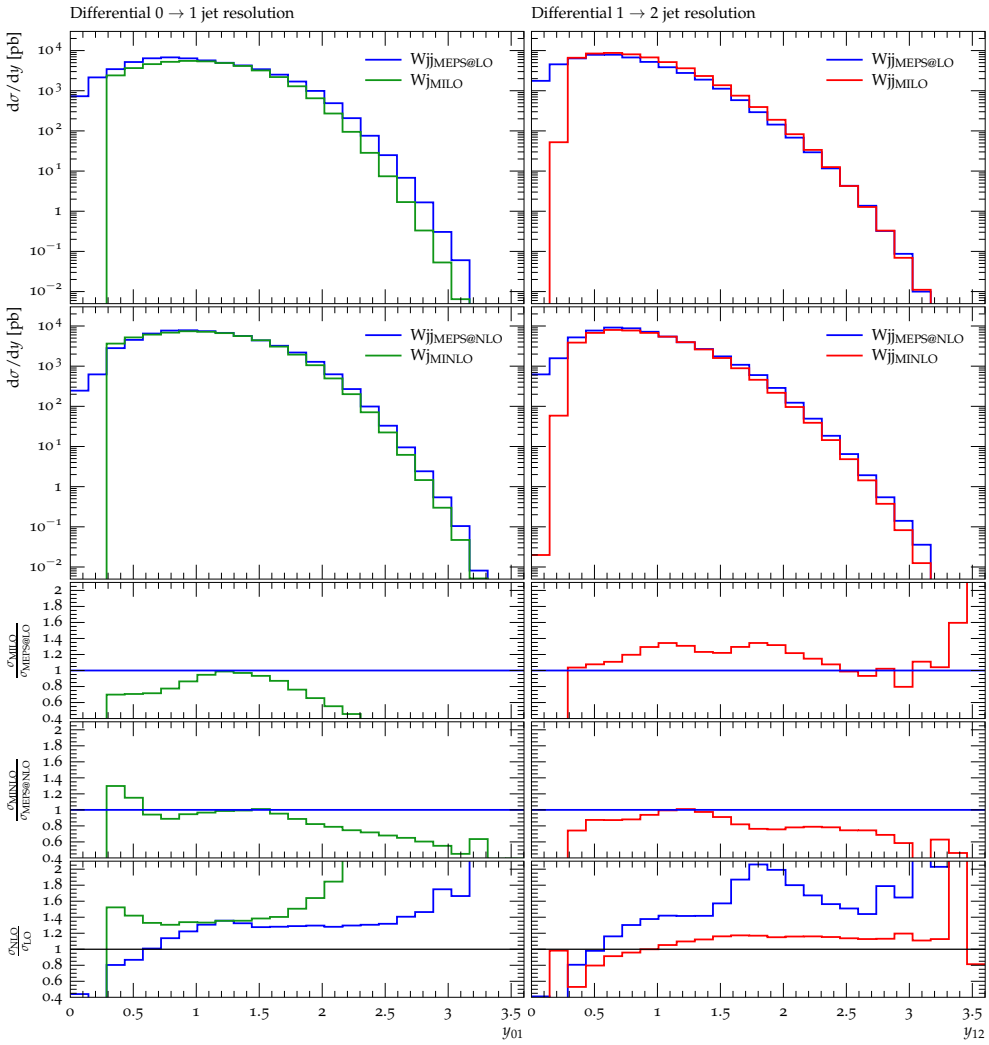
Results

In Fig. 1.3, we start comparing the differential jet rates, defined as

$$y_{i\,i+1} = \log_{10}\left(\frac{d_{i\,i+1}}{1\text{GeV}}\right), \tag{1.53}$$

where $d_{i\,i+1}$ is the k_T measure associated with the clustering from the $i+1$-jet configuration to a i-jet configuration. As for the $0 \to 1$ differential jet rate (Fig. 1.3(a)), we see that all predictions are well behaved in the soft limit, both at leading and next-to-leading order. Such behaviour in the MiNLO algorithm is given by the analytic Sudakov form factors defined in (1.48), which damps the otherwise divergent matrix elements in the soft-collinear region. Note that the Sudakov peak is expected to be at few GeV, close to the employed generation cuts and to the shower cutoff. This might bias the results in the very soft region. At LO, there is a small shape distortion between $Wj_{\texttt{MiLO}}$ and $Wjj_{\texttt{MePs@LO}}$ in the Sudakov region, where the agreement goes from the 20% level in the extremely soft region (where MiLO has a generation cut) to less than 5–10% at $d \sim 30\,\text{GeV}$. In the intermediate and hard regions the two LO predictions differ a lot as expected, with discrepancies of 50% or more. At NLO the agreement improves. The $Wj_{\texttt{MiNLO}}$ simulation agrees within 10% or less with MePs@NLO in the whole soft region, apart from an excess of +20% near the generation cut threshold. We stress that the results at small transverse momentum in MiNLO are driven by the effect of the analytic Sudakov form factors and the NLO subtraction term of (1.49). In the hard tail we observe a discrepancy with respect to MePs@NLO, as expected, which is nevertheless milder than in the LO case. The origin of such behaviours in the hard region could be explained by the fact that the simplest way for a W to emit a hard jet is a configuration where the boson is relatively soft, and the recoil is given by another hard jet. Such topology is not present in $Wj_{\texttt{MiLO}}$, whereas it is in MePS (due to the Wjj MEs) and in $Wj_{\texttt{MiNLO}}$ (due to the real emission). This interpretation is supported by the fact that MePS has a K-factor reasonably close to one across the whole plotted range, while in MiNLO the K-factor is strongly enhanced in the hard region (bottom panel in Fig. 1.3(a)).

The $1 \to 2$ differential jet rate is shown in Fig. 1.3(b). In the $Wj_{\texttt{MiNLO}}$ simulation the second jet is not resummed. Thus only $Wjj_{\texttt{MiNLO}}$ predictions are considered. In the 2-jet case the agreement between the two algorithms is decent at leading order ($\sim 30\%$) in almost all the range, with a mild shape



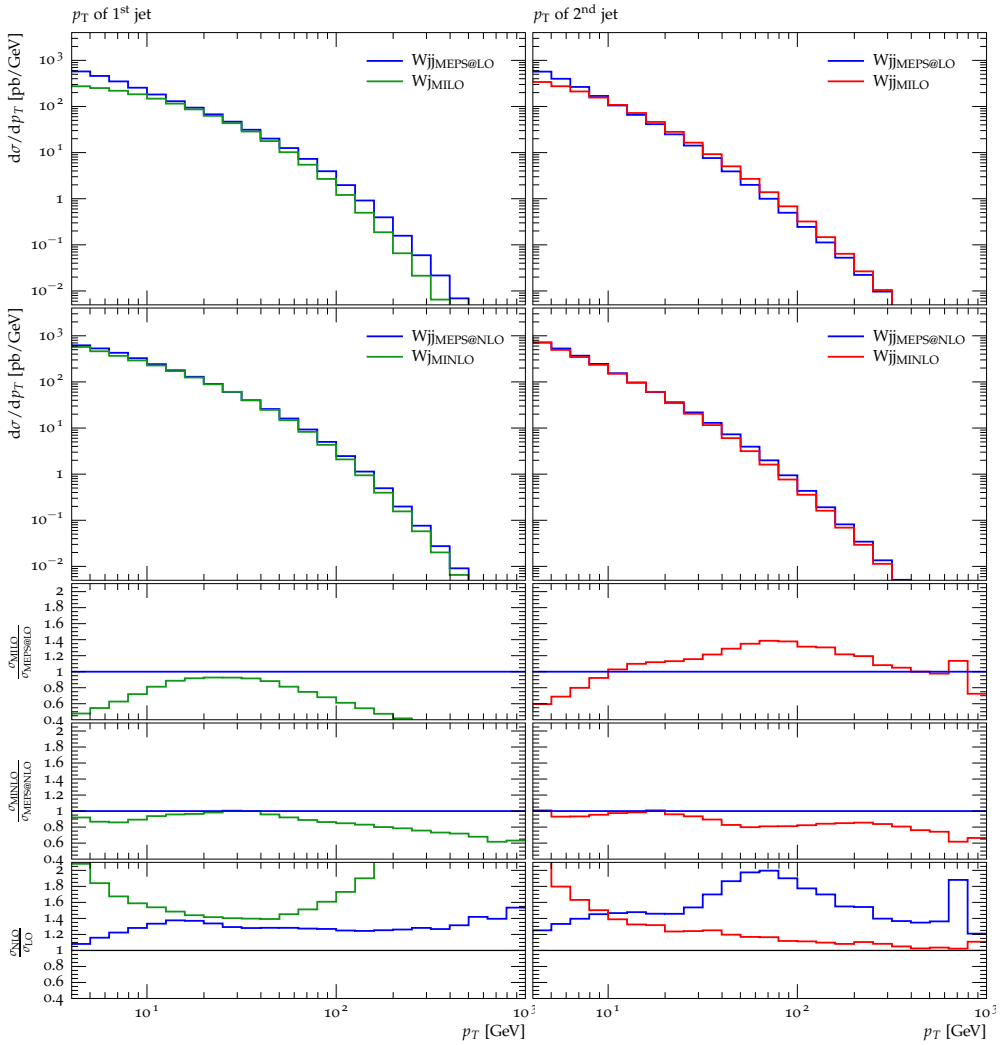
(a) (b)

Figure 1.3: Differential jet rates $y_{i\,i+1}$ associated with the $1 \to 0$ (1) and $2 \to 1$ (b) clustering. The top panels show absolute predictions at LO, whereas the second show the NLO ones. The third and fourth panels show the ratios between the various LO and NLO predictions, respectively. Finally, in the bottom panel the ratio NLO/LO for each prediction is shown.

difference, whereas the NLO predictions turn out to be closer in shape in the soft region. It is interesting to notice that the MiNLO algorithm yields a rather stable and small K–factor in the whole range.

In Fig. 1.4 the p_T of the first and second hardest jets are plotted. The former is equivalent to y_{01}, but shown in different form and range. In the

transverse momentum spectrum of the second hardest jet we observe that the agreement in the Sudakov region improves from 20–30% at LO to order of 10% at NLO. Again, we see that MiLO simulations yield lower predictions than MePs@LO ones in the very soft regime, whereas this discrepancy disappears at NLO.



(a) (b)

Figure 1.4: Transverse momentum spectrum of the 1^{st} (a) and 2^{nd} (b) hardest jet. Ratio plots as in Fig. 1.3.

Summary and conclusions

We presented a comparison between the MEPS algorithm implemented in SHERPA and our MINLO algorithm in case of W+jets production, at both leading and next-to-leading order. The aim of the study was to validate our MINLO implementation against the MEPS algorithm regarding the treatment of Sudakov logarithms. For this reason, we focused our analysis on the very soft region and on jet observables.

Both MILO and MINLO predictions yield well behaved results in the Sudakov region, which is a consequence of the correct exponentiation of large logarithms. Despite the MEPS and MINLO methods differ by aspects that prevent a precise quantitative comparison, we found nice consistency between the predictions in the Sudakov region.

We also observed a better agreement in passing from LO to NLO accuracy, where it goes from 20% and mild shape distortions at leading order to $\sim 10\%$ and small shape distortions at next-to-leading order. The good agreement found in this comparison demonstrates the consistency of both NLO matching and logarithmic resummation in our MINLO implementation.

A systematic comparison of the two methods would need the matching of the MINLO algorithm to the Sherpa parton shower, and we plan to address this aspect in future studies.

In Chapter 3 we report an application of the MINLO algorithm on $t\bar{t}$+1, 2 and 3 jets at NLO QCD.

1.5 Automated tools

In order to compare theoretical predictions to experimental data, computer programs, which simulate high-energy collisions in a generic and widely automated way, have been developed. These programs are known as event generators. Event generators are based on the factorisation of scattering processes into various stages, each with its energy scale. This is depicted in Fig. 1.5. The hardest scale is related to the hard process (dark red blob). Here is where perturbation theory holds and this part of the simulation is described in terms of matrix elements (MEs). Coloured partons then undergo cascades of collinear splitting processes that lead to showers of low energy partons. This is handled by parton showers, which describe the evolution of QCD partons from the hard scale down to hadronisation scale (of order of $\Lambda_{QCD} \sim 1$ GeV). At this stage coloured partons are converted into colourless hadrons, by means of hadronisation models. Unstable hadrons then decay into the particles observed in detectors (dark green blobs). In

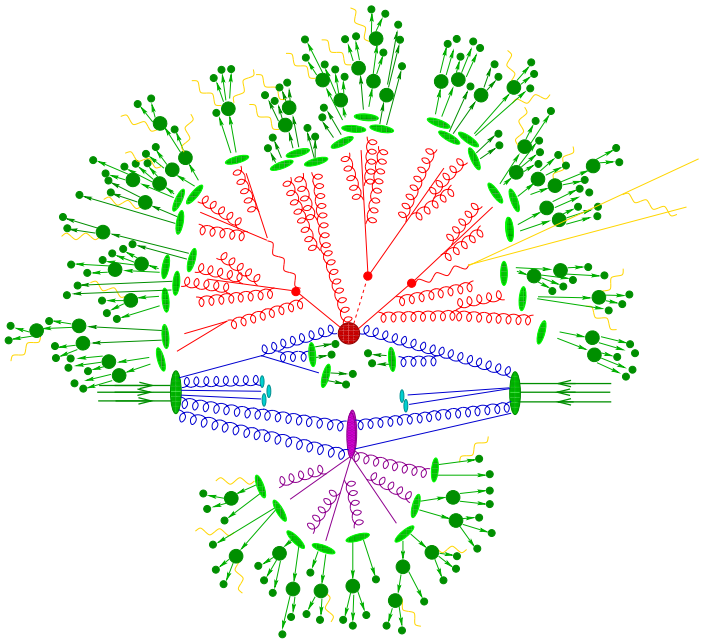


Figure 1.5: Pictorial representation of a generic event as produced by an event generator. The hard interaction (production plus decays) corresponds to the red blobs. Additional hard QCD radiation is produced as a shower of collimated low-energy partons (red) and a secondary interaction takes place (purple blob) before the final-state partons hadronise (light green blobs) and hadrons decay (dark green blobs). Photon radiation occurs at any stage (yellow).

case of hadronic collisions, a secondary interaction can occur between the hadron remnants (purple blob in Fig. (1.5)), which gives rise to additional QCD radiation. This part of the process is called underlying event and is described by means of phenomenological models. Event generators have been developed during the years parallel to new experiments at increasingly higher energies. The increasing complexity of the final states and the need of higher accuracy require these codes to be increasingly fast and flexible. The most prominent examples of event generators are PHYTIA 8 [66], HERWIG 7 [67] and SHERPA [28].

Many important experimental signatures at the LHC are characterised by high complexity. In particular they typically involve a large number of final-state particles. This has lead to the creation and perfection of powerful theoretical techniques and tools, which can deal with multi-partons final states. By now, at tree level, several codes that can compute corresponding cross sections and generate events in a fully automated way are available.

The most prominent examples include AlpGen [68], MG5aMC@NLO [69] and Amegic++ [70]. Amegic++ is part of a fully equipped event generation framework, the Sherpa [28] event generator.

The need of higher accuracy has pushed towards the development of next-to-leading order calculations for the hard processes, and towards more realistic simulations which can be used in experimental contexts. These aims brought to the creation of fast and flexible automated-NLO-amplitudes generators such as OpenLoops [71], NJet [72] and GoSam [73], which can be interfaced to event generators, and to the development of theoretical techniques to combine NLO results with full MC simulations. Examples of such methods are NLO and PS matching and multi-jets merging, some of which have been already introduced in Sections 1.3.1 and 1.3.2.

Among the various MC full event generators we recall MG5aMC@NLO [27] and Sherpa+OpenLoops. The latter is the framework that has been used throughout this work, and it will be introduced in this section.

OpenLoops [71] is a one-loop generator based on a novel numerical recursion, which is formulated in terms of loop-momentum polynomials called "open loops" and allows for a fast evaluation of scattering amplitudes with many external particles. This is natively interfaced to Sherpa [28], a complete event-generation framework, which supports LO matrix elements generation, a subtraction mechanism, matching and merging with an internal parton shower, hadronisation and simulation of multiparton interactions. Thanks to its speed and flexibility, the full framework has been successfully employed in several complicated processes [32, 37, 74–78], and has been used throughout this work.

1.5.1 Sherpa

Sherpa is an acronym for "Simulation of High Energy Reactions of PArticles". It consists of a fully-equipped event generation framework, which has been completely written in the object oriented programming language C++. It includes two automatic matrix-element generators Amegic and Comix [79], a parton shower based on Catani-Seymour dipoles, a multiple parton interaction module, a hadronisation module, a hadron and tau decay library, and a program for the simulation of QED radiation. Amegic is a tree-level element generator based on Feynman diagrams, which are translated to helicity amplitudes following the methods of [80]. It has been thoroughly tested for multiparticle production in several processes [81]. In addition to the Standard Model, it implements also the MSSM [82], the ADD [83] model, the effective ggH theory and other models. Comix is a multi-leg tree-level matrix-element generator, based on the colour-dressed Berends-Giele

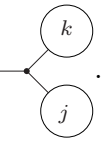
recursive relations [84]. It employs an algorithm to recursively compute phase-space weights. The evaluation of initial- and final-state integrals is carried out by following the adaptive multi-channel methods of [85, 86] together with a Vegas optimisation [87], but also other phase-space generators are available. The Catani-Seymour subtraction [88] method is employed and automated for both matrix-element generators. A parton shower based on Catani–Seymour (CS) dipoles, proposed in [89, 90], has been implemented in SHERPA [53]. It relies on the factorisation of real-emission matrix elements in the CS subtraction framework, where the dipole functions, taken in four dimensions and averaged over spins, are used as shower splitting kernels. When interfaced to an external one-loop generator, the framework allows to match NLO MEs to the parton shower according the S-MC@NLO method, and to consistently combine different final-state multiplicities with the parton shower at NLO accuracy following the MEPS@NLO prescription (see Sections 1.3.1 and 1.3.2). These methods have been implemented within Sherpa in a fully-general and automated way and can be applied to any process within the Standard Model. The underlying event is simulated by AMSIC++ [91], according to the method in [92]. In SHERPA the treatment of multiple interactions has been extended by allowing for the simultaneous evolution of an independent parton shower in each of the subsequent (semi-)hard collisions. The beambeam remnants are organized such that partons which are adjacent in colour space are also adjacent in momentum space. Finally, hadronisation effects and hadron decays are implemented and fully automated. The former are essentially based on the continuation of a dipole-shower model into the non-perturbative regime, where the strong coupling is parametrised and can be tuned to better fit the data. The kinematics of the splittings and hadrons is then chosen according to Lorentz invariant evolution parameters [93]. The hadron decay module employs several matrix elements and form-factor models, and resulting decay products respect full spin correlations. SHERPA itself is the steering module that manages all the above and controls the event generation at all stages. It also implements the initialisation of the generator and the interaction with external interfaced codes.

1.5.2 OpenLoops

In this section the main aspects of the OPENLOOPS algorithm will be outlined following [71, 94]. OPENLOOPS is an automated tree-level and one-loop generator based on a novel numerical recursion, which is formulated in terms of loop-momentum polynomials called "open loops" and allows for a fast evaluation of scattering amplitudes with many external particles. The key

idea behind this algorithm is the recursive construction of the colour-stripped amplitudes in terms of objects called *sub-trees*, starting from external wave functions. At tree level, sub-trees are obtained from tree diagrams after a line is cut and are represented as complex ntuples $w^\beta(i)$, with β being the Lorentz or spinor index of the cut line and i an index which embodies the topology, the (off-shell) momentum and the particle contents, i.e.,

$$w^\beta(i) = \quad = \quad . \tag{1.54}$$

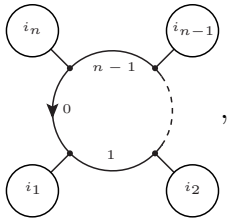


The cut lines are indicated by dots and external lines are not shown[7]. In this framework, the recursive step reads

$$w^\beta(i) = \frac{X^\beta_{\gamma\delta}(i,j,k)\; w^\gamma(j)\; w^\delta(k)}{p_i^2 - m_i^2 + i\varepsilon}, \tag{1.55}$$

where $X^\beta_{\gamma\delta}/(p_i^2 - m_i^2 + i\varepsilon)$ describes a vertex connecting i, j, k, and a propagator attached to i. The recursion ends when all the sub-trees needed to build all the diagrams have been generated. Sub-trees which occur in different diagrams are calculated once and then re-used in all the occurrences. The actual implementation of the code makes use of universal routines built of wave functions, vertices and propagators corresponding to the Feynman rules of the theoretical model at hand. This technique is also used for 1-loop diagram calculation. Each n-points colour-stripped diagram $\delta A^{(d)}$ is interpreted as an ordered set of n sub-trees, $\mathcal{I}_n = \{i_1, \ldots, i_n\}$, connected by loop propagators,

$$\delta A^{(d)} = \int \frac{d^D q\, \mathcal{N}(\mathcal{I}_n; q)}{D_0 D_1 \ldots D_{n-1}} = \quad , \tag{1.56}$$



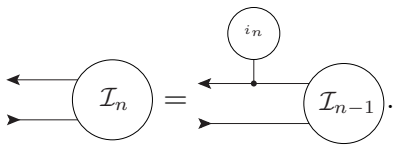
where the denominators $D_i = (q+p_i)^2 - m_i^2 + i\varepsilon$ are functions of the loop momentum q, the external momenta p_i and internal masses m_i. The rest of the contributions coming from vertices, propagators and sub-trees are encoded in the numerator $\mathcal{N}(\mathcal{I}_n; q)$, which is a polynomial in the loop momentum q of degree $R \leq n$,

$$\mathcal{N}(\mathcal{I}_n; q) = \sum_{r=0}^{R} \mathcal{N}_{\mu_1 \ldots \mu_r}(\mathcal{I}_n)\; q^{\mu_1} \ldots q^{\mu_r}. \tag{1.57}$$

[7]For brevity, quartic vertices are omitted, but their inclusion is straightforward.

Setting $p_0 = 0$ and cutting the denominator D_0, the result is a tree level structure, and the numerator function can be built in a recursive way attaching sub-trees,

$$\mathcal{N}_\alpha^\beta(\mathcal{I}_n;q) = \quad = \quad . \tag{1.58}$$



Here α and β are respectively the spinor and Lorentz index of the cut propagator. The numerator $\mathcal{N}_\alpha^\beta(\mathcal{I}_n;q)$ is then built recursively according to (1.55),

$$\mathcal{N}_\alpha^\beta(\mathcal{I}_n;q) = X_{\gamma\delta}^\beta(\mathcal{I}_n,i_n,\mathcal{I}_{n-1})\,\mathcal{N}_\alpha^\gamma(\mathcal{I}_{n-1};q)\,w^\delta(i_n), \tag{1.59}$$

with $X_{\gamma\delta}^\beta$ being the same vertices as in the tree recursion. At this stage, the numerator is split into a loop-momentum-dependent part Z and a independent part Y,

$$\mathcal{N}_\alpha^\beta(\mathcal{I}_n;q) = \sum_{r=0}^{R} \mathcal{N}_{\mu_1\dots\mu_r;\alpha}^\beta(\mathcal{I}_n)\,q^{\mu_1}\dots q^{\mu_r}, \quad X_{\gamma\delta}^\beta = Y_{\gamma\delta}^\beta + Z_{\mu;\gamma\delta}^\beta q^\mu, \tag{1.60}$$

leading to the recursive relation for the so-called *open loops*

$$\mathcal{N}_{\mu_1\dots\mu_r;\alpha}^\beta(\mathcal{I}_n) = \Big[Y_{\gamma\delta}^\beta\,\mathcal{N}_{\mu_1\dots\mu_r;\alpha}^\gamma(\mathcal{I}_{n-1}) + Z_{\mu_1;\gamma\delta}^\beta\,\mathcal{N}_{\mu_2\dots\mu_r;\alpha}^\gamma(\mathcal{I}_{n-1})\Big]\,w^\delta(i_n). \tag{1.61}$$

The recursion stops with the contraction of the α, β indices $\mathcal{N}_{\mu_1\dots\mu_r} = \mathcal{N}_{\mu_1\dots\mu_r;\alpha}^\alpha$, yielding the coefficients of the tensor integral representation of the diagram

$$\delta A^{(d)} = \sum_{r=0}^{R} \mathcal{N}_{\mu_1\dots\mu_r}(\mathcal{I}_n)\,T_{n,r}^{\mu_1\dots\mu_r} \quad \text{with} \quad T_{n,r}^{\mu_1\dots\mu_r} = \int \frac{d^Dq\;q^{\mu_1}\dots q^{\mu_r}}{D_0D_1\dots D_{n-1}}. \tag{1.62}$$

The tensor integrals $T_{n,r}^{\mu_1\dots\mu_r}$ are reduced to m-points scalar integrals $T_{m,0}$ with $m = 1,2,3,4$. Alternatively, this reduction can be avoided through the OPP method [95] by means of a direct connection between the numerator functions $\mathcal{N}(\mathcal{I}_n;q)$ and the scalar integral representation of the amplitude. In this case, the coefficients can be determined by multiple evaluations of $\mathcal{N}(\mathcal{I}_n;q)$ for loop momenta q constrained by multiple-cut conditions of the form $D_i = D_j = \dots = 0$. The described algorithm is implemented in the OpenLoops program [96]. Feynman diagrams are generated by Feynarts [97], whereas Mathematica manages the recursive routines and generates the related FORTRAN90 codes. The generated code is organised in libraries

which can be loaded an called by means of the native FORTRAN90/C++ interface. The tensor reduction is dealt with by COLLIER [98–101] or CutTools [102] and OneLOop [103]. Several options are settable to guarantee numerical stability. The code results to be light, fast and stable, which are crucial features for multi-final states Monte Carlo simulations. OPENLOOPS has been successfully employed in several QCD and EW processes [32, 37, 74–76, 78, 104–109], and, in conjunction with SHERPA, it has been used throughout this work.

Part II

Phenomenological applications

Introduction

The $t\bar{t}H$ channel permits the direct measurement of the top Yukawa coupling. In the dominant Higgs-decay mode, $H \to b\bar{b}$, this also allows to constrain the bottom Yukawa coupling. However, the huge QCD background to $t\bar{t}H(b\bar{b})$ production, which consists of $t\bar{t}$+light jets production and $t\bar{t}b\bar{b}$ production, renders $t\bar{t}H(b\bar{b})$ searches very challenging. The signal over background ratio, indeed, is typically at the percent level and hardly reaches ten percent, depending on the signal region [13, 14]. The ambiguity in the Higgs mass peak identification due to the presence of four b quarks in the final state coming from Higgs and top decays represents another obstacle in this search. In this context, the availability of precise and realistic theoretical predictions for the relevant channels is of paramount importance to reduce the systematic uncertainties associated with the experimental analyses. The theoretical predictions for these backgrounds are particularly challenging mainly for two reasons. First, they involve multi-particle final states with six or more coloured legs, two of which are heavy flavours, which makes high order calculations demanding. Second, the multi-scale nature of the processes, with masses and kinematic invariants that range from $m_b \sim 5\,\text{GeV}$ to $m_{t\bar{t}} \sim 500$ GeV, renders the assessment of theoretical uncertainties and the matching NLO matrix elements to parton showers nontrivial.

Fixed-order calculations at NLO QCD accuracy can reduce the theoretical uncertainties of the backgrounds to 10–30% [29, 30, 33–35], whereas more recent NLO predictions combined to parton showers yield more realistic results, thus being suitable for experimental analyses. The MEPS@NLO simulation of top-pair production in association with up to two light jets, published in [32], currently represents the state of the art in multijet merging at NLO. Regarding the irreducible background to the $t\bar{t}H(b\bar{b})$ signal, NLO+PS results for $t\bar{t}b\bar{b}$ production using the POWHEG method and the S-MC@NLO method have been published in [36] and [37], respectively. While the former calculation handles b-quarks as massless partons, the latter one—reported in this chapter— includes b-mass effects and represents the first application of NLO matching to a process with four coloured massive particles in the final

state. The combination of the small theoretical uncertainties of NLO calculations with the more physical description of the final states by the parton shower makes these simulations particularly suitable for the applications to experimental searches. Let us note in passing that fully realistic Monte Carlo simulations at NLO accuracy of signal and backgrounds, based on the results presented in this work, have been employed to study the potential sensitivity of the LHC in the $t\bar{t}H(b\bar{b})$ channel [77].

This part of the manuscript is devoted to applications of the theoretical methods and automated tools presented in Part I in the context of $t\bar{t}H(b\bar{b})$ searches. We report two works, based on [37] and [39], concerning the $t\bar{t}$+multijets and $t\bar{t}b\bar{b}$ processes, which constitute the dominant backgrounds to the $t\bar{t}H(b\bar{b})$ signal.

In Chapter 2 we present a next-to-leading order simulation of $t\bar{t}b\bar{b}$ production with massive b-quarks matched to the Sherpa parton shower. The inclusion of the b–mass effects allows to extend NLO predictions to arbitrary $t\bar{t}b\bar{b}$ kinematics, including the case where one or both b-jets arise from collinear $g \to b\bar{b}$ splittings. We find that this splitting mechanism plays an important role for $t\bar{t}H$ searches at the LHC. In section 2.1 we report the work published in [37], whereas the rest of the chapter is focused on additional unpublished studies regarding $t\bar{t}b\bar{b}$ production. In particular, in section 2.2 we show a more extensive study, which has not been included in the original work. Finally, in section 2.3 we present a technical study on the resummation scale dependence of the Mc@NLO and S-Mc@NLO methods implemented in the MG5aMC@NLO+Phytia 8 and Sherpa+OpenLoops frameworks respectively, applied to the $t\bar{t}b\bar{b}$ simulation at 13 TeV.

In Chapter 3 we present theoretical predictions for the production of top-quark pairs with up to three light jets at the next-to leading order in perturbative QCD. The relevant calculations are performed with Sherpa and OpenLoops. To address the issue of scale choices and related uncertainties in the presence of multiple scales, we compare results obtained with the standard scale $H_{\mathrm{T}}/2$ at fixed order and the MiNLO procedure. Analysing various cross sections and distributions for $t\bar{t} + 0, 1, 2, 3$ jets at the 13 TeV LHC we find a remarkable overall agreement between fixed-order and MiNLO results. The differences are typically below the respective factor-two scale variations, suggesting that for all considered jet multiplicities missing higher-order effects should not exceed the ten percent level.

Chapter 2

Precise predictions for ttbb production at the LHC

2.1 NLO+PS predictions for ttbb production at the 8TeV LHC

The recent discovery of the Higgs boson and first measurements of its interactions permit to probe the mechanism of spontaneous symmetry breaking, by which elementary particles acquire their mass [1, 2]. Data collected in the first run of the LHC provide significant sensitivity to Higgs-boson interactions with force carriers—gluons, photons, Z and W bosons—while constraints on Higgs-couplings to matter particles—leptons and quarks—are less stringent and mostly stemming from indirect effects on Higgs–gluon and Higgs–photon couplings. The direct investigation of Higgs-boson couplings to quarks and leptons will thus represent a crucial further step towards a complete understanding of the origin of mass. In this context, the reaction pp $\to t\bar{t}H(b\bar{b})$, i.e. Higgs-boson production in association with a top-quark pair with subsequent Higgs-boson decay into a bottom-quark pair, provides a unique opportunity to test the mass-generation mechanism in the heavy-quark sector. This process is notoriously very challenging due to the presence of four b-quarks in the final state, which hampers a correct identification of the Higgs-boson mass peak. As a result, the $t\bar{t}H$ signal is strongly contaminated by background contributions from top-quark pair production in association with light-, charm- and bottom-jet pairs. The large uncertainty in the Monte-Carlo simulations of these multi-particle QCD backgrounds represents one of the main bottlenecks of the present $t\bar{t}H(b\bar{b})$ analyses [110, 111], and the availability of state-of-the art theory predictions for $t\bar{t}jj$, $t\bar{t}c\bar{c}$, and $t\bar{t}b\bar{b}$ production is a key prerequisite to improve the sensitivity to the $t\bar{t}H(b\bar{b})$ sig-

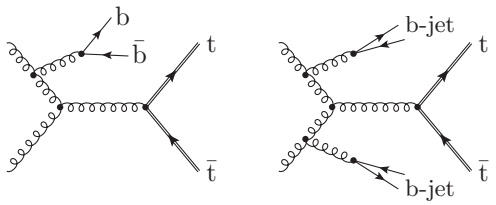


Figure 2.1: Tree topologies corresponding to $t\bar{t}b\bar{b}$ production via single hard (left) or double collinear (right) $g \to b\bar{b}$ splitting.

nal. In the case of the irreducible $t\bar{t}b\bar{b}$ background, theory predictions play an especially important role, since the lack of sufficiently distinctive kinematic features and the rather small cross section do not allow for an efficient $t\bar{t}b\bar{b}$ measurement in a signal-free control region.

NLO calculations for $t\bar{t}b\bar{b}$ [33–35, 112] and $t\bar{t}jj$ [29, 30] production can reduce perturbative uncertainties from 70–80% down to 15–20%. However, in order to be applicable to the experimental analyses, these calculations need to be matched to parton showers. Matched NLO predictions for $pp \to t\bar{t}+ \leq 1$ jets, with consistent merging of 0- and 1-jet final states, have been presented in [31], and first technical results towards NLO matched $t\bar{t}b\bar{b}$ production have been discussed in [36], where the NLO calculation of [35] was matched at the level of the first shower emission with the PowHeg approach [17]. In this letter, we present a fully-showered NLO simulation of $t\bar{t}b\bar{b}$ production. Besides matching NLO matrix elements to the parton shower with the MC@NLO method [15], for the first time we also include finite b-quark mass effects. This represents the first complete NLO-matched simulation with four (massive) coloured particles in the final state. Using massive b-quarks we can extend the simulation to the whole $t\bar{t}b\bar{b}$ phase space, thereby including also $t\bar{t}+1$ b-jet contributions with an unresolved (soft or collinear) b-quark, which play an important role in the $t\bar{t}\mathrm{H}(b\bar{b})$ analysis. Moreover, matching massive NLO matrix elements to the parton shower gives access to novel $t\bar{t}$ + b-jets production mechanisms, where b-jets arise from hard gluons via collinear $g \to b\bar{b}$ splittings. In particular, one can describe $t\bar{t}$ + 2 b-jet events where both b-jets originate from $g \to b\bar{b}$ splittings (see Fig. 2.1). For this kind of configurations—which turn out to be quite important—the finite b-quark mass allows one to obtain an NLO accurate description of the first $g \to b\bar{b}$ splitting, while simulations with massless b-quarks must rely on $t\bar{t}$gg matrix elements plus pure parton-shower splittings in the collinear regions.

The presented simulation has been prepared within the Sherpa +OpenLoops framework [28, 71, 74], which supports the fully automated simulation of any Standard-Model process at NLO QCD, including matching to the par-

ton shower and multi-jet merging. The OpenLoops [71] program is a one-loop generator based on a novel numerical recursion, which is formulated in terms of loop-momentum polynomials called "open loops" and allows for a fast evaluation of scattering amplitudes with many external particles.[1] It uses the Collier library [114] for the numerically stable evaluation of tensor integrals [98,99] and scalar integrals [100]. Real-emission contributions, infrared subtractions based on the Catani–Seymour (CS) technique [46, 115], and phase-space integration are handled by Sherpa [28] and Amegic++ [70]. The NLO corrections are matched to the Sherpa parton shower [53] using the Sherpa formulation [18, 116] of the MC@NLO method [15] (S-Mc@NLO).[2] The essence of the S-Mc@NLO approach is encoded in the following formula for the no-emission and first-emission contributions to the expectation value of a generic observable [116],

$$\begin{aligned}\langle\mathcal{O}\rangle = &\int \mathrm{d}\Phi_B \Big[B(\Phi_B)+V(\Phi_B)+I(\Phi_B)\Big]U(t_0,\mu_Q^2)\\ &+\int \mathrm{d}\Phi_R \Big[R(\Phi_R)-\sum_{ijk} D_{ijk}(\Phi_R)\theta(\mu_Q^2-t)\Big]\mathcal{O}(\Phi_R).\end{aligned} \tag{2.1}$$

The terms $B(\Phi_B)$ and $V(\Phi_B)$ represent Born and virtual matrix-element contributions to the Born phase space Φ_B, while $R(\Phi_R)$ denotes real-emission matrix-element contributions to the corresponding phase space Φ_R. Similarly as for NLO calculations, infrared singularities are removed from the Φ_R phase space via local subtraction terms $D_{ijk}(\Phi_R)$ and added back to the virtual contributions in the form

$$I(\Phi_B) \;=\; \sum_{ijk}\int \mathrm{d}\Phi_{R|B} D_{ijk}(\Phi_R)\theta(\mu_Q^2-t), \tag{2.2}$$

where each subtraction term is integrated over a factorised phase space $\Phi_{R|B}$ associated with a $\Phi_R \to \Phi_B$ mapping. In fixed-order calculations, to achieve an exact cancellation of the subtraction terms, events associated with $D_{ijk}(\Phi_R)$ must be attributed to the Born phase space according to the appropriate $\Phi_R \to \Phi_B$ mapping. In contrast, in the S-Mc@NLO approach $D_{ijk}(\Phi_R)$ contributions are handled as genuine real-emission events, and the resulting mismatch of the form $D_{ijk}(\Phi_R)\,[\mathcal{O}(\Phi_R)-\mathcal{O}(\Phi_B)]$ is compensated, to order α_s, by $\Phi_B \to \Phi_R$ migrations that result from parton-shower emis-

[1] A public implementation of OpenLoops will appear in the next future [113].

[2] In the following, S-Mc@NLO always refers to the algorithm of Refs. [18, 116] and its implementation within Sherpa.

sions. The first shower emission is described by

$$U(t_0,\mu_Q^2) = \Delta(t_0,\mu_Q^2)\mathcal{O}(\Phi_B) + \sum_{ijk}\int_{t_0}^{\mu_Q^2} \mathrm{d}\Phi_{R|B}\frac{D_{ijk}(\Phi_R)}{B(\Phi_B)}\Delta(t,\mu_Q^2)\mathcal{O}(\Phi_R), \tag{2.3}$$

where the second line corresponds to the first-emission probability, and the Sudakov form factor $\Delta(t_0,\mu_Q^2)$ represents its no-emission counterpart. The parton shower is driven by the evolution variable t. It starts at the resummation scale μ_Q^2 and stops when t reaches the infrared cut-off t_0. The key principle, by means of which the S-MC@NLO approach preserves NLO accuracy up to the first emission, is the correspondence between the splitting kernels of the parton shower and the terms D_{ijk} that are subtracted from the real emission. In SHERPA this is achieved by using CS dipoles D_{ijk} both as subtraction terms and as splitting kernels of the parton shower. More precisely, the kernels of the shower are given by the spin-averaged CS dipoles, taken in the large-N_c limit. In addition, to obtain a fully consistent matching, the first shower emission is supplemented by exact spin and colour correlations [18]. The S-MC@NLO matching can be regarded as an effective subtraction of the first shower emission, and, similarly as for the shower, also the subtraction terms in (2.1) and (2.2) must be restricted to the kinematic region $t < \mu_Q^2$. Finally, no-emission and first-emission events generated according to (2.1)–(2.3) are used as seeds for subsequent shower emissions.

In the following, we present and compare LO, NLO and S-MC@NLO simulations of $t\bar{t}b\bar{b}$ production at the 8 TeV LHC. The results are based on a SHERPA 2.0 pre-release version.[3] Hadronisation and underlying events are not considered, and top quarks are treated as stable particles with mass $m_t = 173.2$ GeV. While spin-correlated $t \to Wb$ decays can be simulated in a fully automated way, omitting top decays permits us to focus on the behaviour of those b-jets that arise from QCD interactions, and that involve many more subtleties from the viewpoint of the theoretical simulation and its uncertainties. Consistently with the use of a finite b-quark mass, $m_b = 4.75$ GeV, we employ four-flavour parton distributions. Specifically, at NLO (LO) QCD the LHApdf implementation of the MSTW2008NLO (LO) parton distributions [117] and the corresponding α_s values are used. While the four-flavour running of α_s misses top- and bottom-quark loop effects, corresponding $\mathcal{O}(\alpha_s)$ contributions are consistently included in the virtual corrections via zero-momentum subtraction of the heavy-quark loops in the renormalisation of α_s.

[3] This version corresponds to SVN revision 23546, which implements a recent tune of the SHERPA parton shower to LEP data.

	ttb	ttbb	ttbb($m_{\rm bb} > 100$)
$\sigma_{\rm LO}$[fb]	$2644^{+71\%+14\%}_{-38\%-11\%}$	$463.3^{+66\%+15\%}_{-36\%-12\%}$	$123.4^{+63\%+17\%}_{-35\%-13\%}$
$\sigma_{\rm NLO}$[fb]	$3296^{+34\%+5.6\%}_{-25\%-4.2\%}$	$560^{+29\%+5.4\%}_{-24\%-4.8\%}$	$141.8^{+26\%+6.5\%}_{-22\%-4.6\%}$
$\sigma_{\rm NLO}/\sigma_{\rm LO}$	1.25	1.21	1.15
$\sigma_{\rm MC}$[fb]	$3313^{+32\%+3.9\%}_{-25\%-2.9\%}$	$600^{+24\%+2.0\%}_{-22\%-2.1\%}$	$181.0^{+20\%+8.1\%}_{-20\%-6.0\%}$
$\sigma_{\rm MC}/\sigma_{\rm NLO}$	1.01	1.07	1.28
$\sigma^{\rm 2b}_{\rm MC}$[fb]	3299	552	146
$\sigma^{\rm 2b}_{\rm MC}/\sigma_{\rm NLO}$	1.00	0.99	1.03

Table 2.1: Cross sections with standard ttb and ttbb cuts and with an additional cut, $m_{\rm bb} > 100$ GeV. Full S-MC@NLO predictions ($\sigma_{\rm MC}$) are compared to results obtained with parton-shower $g \to b\bar{b}$ splittings switched off ($\sigma^{\rm 2b}_{\rm MC}$). The first and second uncertainty represent $\xi_{\rm R}$ and $\xi_{\rm F}$ variations. In the S-MC@NLO case, the latter is combined with $\xi_{\rm Q}$ variations in quadrature.

As renormalisation scale we employ the geometric average of the top-quark and b-quark transverse energies,[4]

$$\mu_R^4 = \xi_{\rm R}^4 \prod_{i={\rm t},\bar{\rm t},{\rm b},\bar{\rm b}} E_{{\rm T},i} = \xi_{\rm R}^4 \prod_{i={\rm t},\bar{\rm t},{\rm b},\bar{\rm b}} \sqrt{m_i^2 + p_{{\rm T},i}^2}\,, \tag{2.4}$$

which represents a natural generalisation of the dynamical scale $\mu^2 = m_{\rm t}\sqrt{p_{\rm T,b} p_{\rm T,\bar{b}}}$ introduced in [34]. The default scale corresponds to $\xi_{\rm R} = 1$, and $\xi_{\rm R}$ parametrises scale variations. To NLO accuracy, this choice corresponds to $\alpha_{\rm s}^4(\mu_R) \simeq \prod_i \alpha_{\rm s}(E_{{\rm T},i})$ and guarantees that the strong-coupling factors associated to the production of the various final-state objects adapt to the respective transverse energies. The factorisation and resummation scales, which define the available phase space for QCD radiation, are related to the average top-quark transverse energy via

$$\mu_F = \frac{\xi_{\rm F}}{2}(E_{\rm T,t} + E_{\rm T,\bar{t}}), \qquad \mu_Q = \xi_Q \mu_F. \tag{2.5}$$

The default scale choice corresponds to $\xi_{\rm F} = \xi_Q = 1$, and $\xi_{\rm F}$ parametrises correlated variations of μ_F and μ_Q, while ξ_Q controls additional variations

[4] Note that a dynamical QCD scale defined in terms of b-quark momenta is infrared safe for $m_{\rm b} > 0$, while for massless b-quarks a scale based on b-jet momenta should be used.

of μ_Q with fixed μ_F. QCD partons, including b-quarks and excluding only top-quarks, are recombined into IR-safe jets using the anti-$k_{\rm T}$ algorithm [65] with jet-resolution parameter $R = 0.4$. Events are categorised according to the number $N_{\rm b}$ of reconstructed b-jets with $p_{\rm T} > 25$ GeV and $|\eta_{\rm b}| < 2.5$. In this respect, we classify as b-jet any jet involving at least a b-quark, which includes also the case of collimated $b\bar{b}$ pairs resulting from the splitting of energetic gluons. This is, at least experimentally, the most realistic b-jet definition, and its implementation at NLO is possible only in presence of massive b-quarks. In fact, in calculations with massless b-quarks, collimated $b\bar{b}$ pairs must be handled as gluon-jets in order to avoid collinear singularities.

To investigate NLO and S-MC@NLO correction effects we considered an exclusive ttbb sample, with events involving $N_{\rm b} \geq 2$ b-jets, and a more inclusive ttb sample with $N_{\rm b} \geq 1$. For the ttbb sample an additional analysis is performed with a cut on the invariant mass of the first and second b-jet, $m_{\rm bb} > 100$ GeV, which corresponds to the $t\bar{t}{\rm H}(b\bar{b})$ signal region. The respective LO, NLO and S-MC@NLO cross sections are reported in Table 2.1. In order to isolate contributions arising from b-quarks emitted by the parton shower, we also present S-MC@NLO predictions generated in absence of ${\rm g} \to b\bar{b}$ parton-shower splittings. Scale uncertainties are assessed via independent factor-two variations of $\xi_{\rm R}$ and $\xi_{\rm F}$. Additional scale uncertainties related to the parton shower are included via $\xi_Q = 2^{\pm 1/2}$ variations of the resummation scale and are combined in quadrature with $\xi_{\rm F}$ variations.

Fixed-order results in Table 2.1 feature NLO K-factors close to 1.2, with ± 0.05 variations depending on the selection cuts. This is consistent with the $\mathcal{O}(20\%)$ contribution of b-quarks to the running of $\alpha_{\rm s}^4(\mu)$ from $m_{\rm b}$ to $\mu_{\rm R}$, and with the fact that the corresponding K-factor in the five-flavour scheme, where b-quark contributions are included in the running of $\alpha_{\rm s}$, is very close to one [118]. In this respect, let us note that a fully consistent resummation of $\ln(\mu_{\rm R}/m_{\rm b})$ terms associated with the running of $\alpha_{\rm s}$ would increase the $t\bar{t}b\bar{b}$ NLO cross section by about 9% as compared to standard 4F-scheme predictions presented in this letter. This estimate was obtained using a modified set of MSTW four-flavour PDFs with five active flavours in the evolution of $\alpha_{\rm s}$.

Scale uncertainties in Table 2.1 are dominated by renormalisation-scale variations and decrease from about 60–70% at LO to 20–30% at NLO. Scale variations at NLO and S-MC@NLO level are rather similar. In presence of standard ttb and ttbb cuts, matching to the parton shower shifts the NLO cross section by only 1% and 6%, respectively. However, the S-MC@NLO correction to $t\bar{t}b\bar{b}$ finals states is quite sensitive to the invariant mass of the $b\bar{b}$ pair and turns out to be enhanced by a factor four in the region $m_{b\bar{b}} > 100$GeV, which is relevant for Higgs-boson searches. This S-MC@NLO effect—which

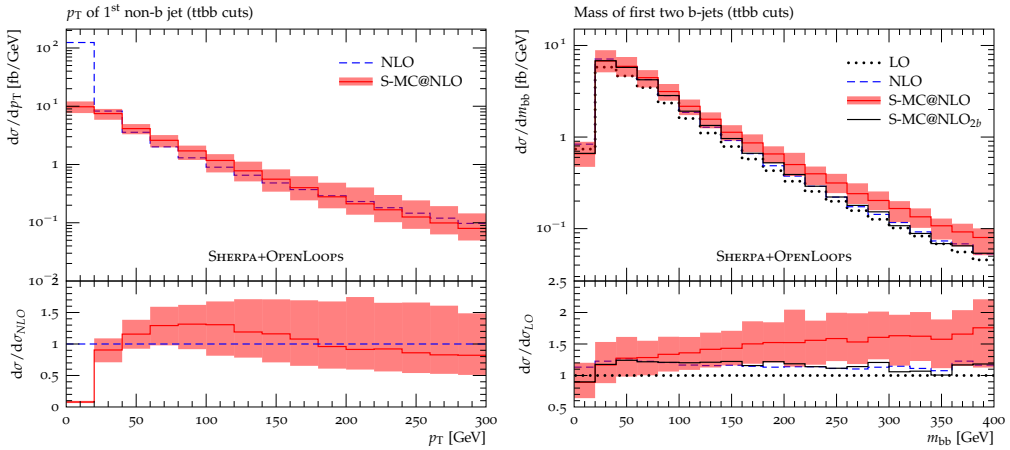


Figure 2.2: Transverse momentum of the first light jet and invariant mass of the first two b-jets with standard ttbb cuts. The S-MC@NLO bands display the combination in quadrature of μ_R, μ_F and μ_Q scale variations. The S-MC@NLO $_{2b}$ curve is obtained by switching off $g \to b\bar{b}$ splittings in the parton shower.

clearly exceeds the magnitude of the Higgs signal in the present $t\bar{t}H(b\bar{b})$ analyses [110, 111]—tends to disappear if $g \to b\bar{b}$ splittings are switched off in the parton shower.[5] As discussed below, various features indicate that this effect is dominated by the double-splitting mechanism depicted in Fig. 2.1.b.

The differential distributions in Figs. 2.2 and 2.3 provide examples of nontrivial matching corrections. Standard ttbb cuts are applied, and the S-MC@NLO bands display the combination in quadrature of μ_R, μ_F and μ_Q scale variations. The corresponding uncertainties are typically around 30% and tend to increase in the tails, also due to statistical fluctuations. The transverse momentum of the first non-b jet (Fig. 2.2.a) shows the typical S-MC@NLO behaviour. At transverse momenta above the resummation scale, where the parton shower stops emitting, S-MC@NLO and NLO predictions agree well. The fixed-order infrared singularity at small p_T is consistently damped by the Sudakov form factor, and Sudakov effects start to be important already at $p_T \sim 50$ GeV. This reflects the presence of intense QCD radiation resulting from the gluon-gluon initial state and from the high center-of-mass energy of the $t\bar{t}b\bar{b}$ system. In the intermediate p_T region we observe an S-MC@NLO correction of about +30% wrt. NLO. This can be

[5] Note that only full S-MC@NLO predictions should be regarded as physical, while results without $g \to b\bar{b}$ parton-shower splittings are showed only for technical aims, namely to illustrate the relevance of multiple $b\bar{b}$ production.

attributed to g $\to b\bar{b}$ parton-shower splittings and to the enhancement of the first shower emission that results from the $(B + V + I)$ term in (2.1). The precise position and magnitude of the S-MC@NLO/NLO maximum depend on the choice of the renormalisation and resummation scales, and scale variations permit assessing related higher-order uncertainties inherent in the matching procedure.

Figure 2.2.b confirms that matching corrections are quite sensitive to the invariant mass of the first two b-jets. The S-MC@NLO/NLO ratio grows with $m_{\rm bb}$ and reaches 25–30% in the Higgs-signal region, $m_{\rm bb} \sim 125$ GeV. This enhancement at high invariant mass can be attributed to $t\bar{t} + 2$ b-jets production via double g $\to b\bar{b}$ splittings, since this mechanism is kinematically favoured by the fact that the probability that two hard gluons split into collinear $b\bar{b}$ pairs does not decrease when the invariant mass of the gluon pair grows. This interpretation is confirmed by the fact that the shape of the S-MC@NLO $m_{\rm bb}$ distribution becomes almost identical to the NLO one if g $\to b\bar{b}$ splittings are switched off in the parton shower. Further evidence of the correctness of the above picture is provided by the fact that the S-MC@NLO excess increases with the di-jet invariant mass at a similar rate as the ratio of the $t\bar{t}$gg to $t\bar{t}b\bar{b}$ cross sections. For instance, using LO matrix elements, we checked that both quantities increase by a factor two in the range between 100 and 250 GeV.

The plots in Fig. 2.3, where an additional cut $m_{\rm bb} > 100$ GeV is applied, reveal distinctive kinematic features of the S-MC@NLO enhancement in the Higgs-signal region. The unambiguous S-MC@NLO/NLO peaks that appear in the distributions, both in the transverse momentum of the first b-jet (Fig. 2.3.a) and in the ΔR separation of the first two b-jets (Fig. 2.3.b), show that the S-MC@NLO enhancement is dominated by back-to-back b-jets with the smallest possible $p_{\rm T}$ that is needed to reach $m_{\rm bb} = 100$ GeV. This is consistent with the expected behaviour of double g $\to b\bar{b}$ splitting contributions in Fig. 2.1.b, where emissions at small-$p_{\rm T}$ are doubly enhanced by soft and collinear singularities associated with the parent gluons. Also this interpretation is fully confirmed by the fact that S-MC@NLO-induced shape distortions in Fig. 2.3 disappear almost completely when g $\to b\bar{b}$ shower splittings are switched off.

To exclude the possibility that double splittings in our simulation are artificially enhanced by a too high choice of the resummation scale, we checked that the characteristic "double-splitting" enhancement in the $m_{b\bar{b}}$ distribution of Fig. 2.2 is present also in simulations based on merged LO matrix elements for $t\bar{t}$ plus multi-jet production. In this framework, $t\bar{t}b\bar{b}$ events are not showered with a global resummation scale, but starting from a scale that is determined according to the most likely shower history of the event

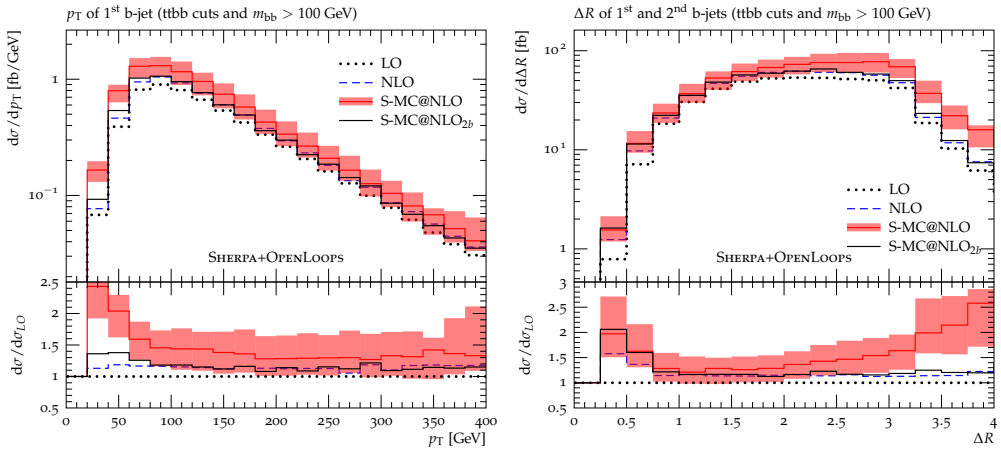


Figure 2.3: Transverse momentum of the first b-jet and ΔR separation of the first two b-jets with standard ttbb cuts and $M_{\rm bb} > 100$ GeV. The S-MC@NLO bands display the combination in quadrature of μ_R, μ_F and μ_Q scale variations. The S-MC@NLO $_{2b}$ curve is obtained by switching off g $\to b\bar{b}$ splittings in the parton shower.

at hand. Comparing the shape of the S-MC@NLO distribution of Fig. 2.2 against MEPS@LO simulations [26] of $t\bar{t}+ \leq 3j$ with massive b-quarks, we found good agreement for merging scales around 15 GeV, i.e. for the case where most of the phase space associated with (the first) g $\to b\bar{b}$ splittings is described in terms of matrix elements, as in the present S-MC@NLO simulation. A thorough understanding of the uncertainties related to the choice of the merging scale and the interplay between matrix elements and parton shower in the vicinity of the kinematic threshold for g $\to b\bar{b}$ splittings requires further detailed studies that are beyond the scope of this letter.

In summary, we presented the first complete S-MC@NLO simulation of $t\bar{t}b\bar{b}$ production at the LHC, including b-quark mass effects. This allows one to cover the full $t\bar{t}b\bar{b}$ phase space at NLO accuracy and to describe contributions stemming from double collinear g $\to b\bar{b}$ splittings, which can lead to a significant contamination of the $t\bar{t}$H($b\bar{b}$) signal. This unexpected finding changes the standard perturbative picture of $t\bar{t}b\bar{b}$ production based on hard b-quark jets. The presented simulation will allow for a thorough analysis of the related uncertainties. In this respect it will be important to assess the role of the parton-shower tune and to devise efficient strategies for the rejection of double-splitting contributions. Aspects not discussed here, such as top-quark decays, hadronisation and underlying events, can be simulated in a fully automated way using SHERPA. To gain more insights into

theoretical uncertainties associated with the parton shower and the b-quark mass, it will be very instructive to compare the four-flavour scheme adopted in this paper to the five-flavour scheme. Both schemes provide reliable NLO predictions for observables involving resolved b-jets at the LHC [119]. In the five-flavour scheme, where b-quarks are massless, $t\bar{t}b\bar{b}$ matrix elements cannot be used to fill the entire b-quark phase space, and the collinear regions need to be described by lower-multiplicity hard matrix elements ($t\bar{t}$g, $t\bar{t}b$, $t\bar{t}$, etc.) supplemented by parton-shower emissions. Technically this requires the merging of NLO matrix elements for $t\bar{t}+0,1,2$ jets, which was presented for the first time in [32]. A consistent combination of this recent simulation and the massive $t\bar{t}b\bar{b}$ predictions presented in this paper would provide an optimal description of $t\bar{t}$ plus multi-jet production.

In the following sections we will present additional studies on $pp \to t\bar{t}b\bar{b}$ that have not been included in [37]. In the first part further observables of the same simulation are showed. The second part is dedicated to an unpublished in-depth investigation of the uncertainties of such simulation, where shape distortions of key observables due to different kinematical dependence of the scales are examined. This study led to the generation of publicly available samples which have been used for $t\bar{t}H(b\bar{b})$ searches by ATLAS [13]. Finally, in the last part we present a technical study on the resummation scale dependence of the S-MC@NLO simulation of $t\bar{t}b\bar{b}$ production at 13 TeV.

2.2 Further observables and theoretical uncertainties at 8 TeV

2.2.1 Extended analysis of $t\bar{t}b\bar{b}$ production

In this section we present further observables which have not been included in [37]. All predictions refer to $pp \to t\bar{t}b\bar{b}$ at 8 TeV and are based on the same setup as described in the previous section. We start showing kinematic observables related to the top quarks. In Fig. 2.4 the p_T and the rapidity of the hardest top are shown. In the $t\bar{t}b$ sample, the NLO fixed order and S-MC@NLO results overlap in the whole range, reflecting the small impact of the shower in this region, whereas with standard $t\bar{t}b\bar{b}$ cuts, S-MC@NLO predictions show a small 10% excess without evident shape distortions. The situation changes slightly when one looks at the transverse momentum of the $t\bar{t}$ pair, displayed in Fig. (2.5). This observable is particularly sensitive to additional QCD emissions, and thus S-MC@NLO results feature a $10-20\%$ ($20\%-30\%$) excess with respect to NLO ones in the mid-high p_T region in the $t\bar{t}b$ ($t\bar{t}b\bar{b}$) sample. This picture is consistent also with b-jets observables

presented in Fig. (2.6).

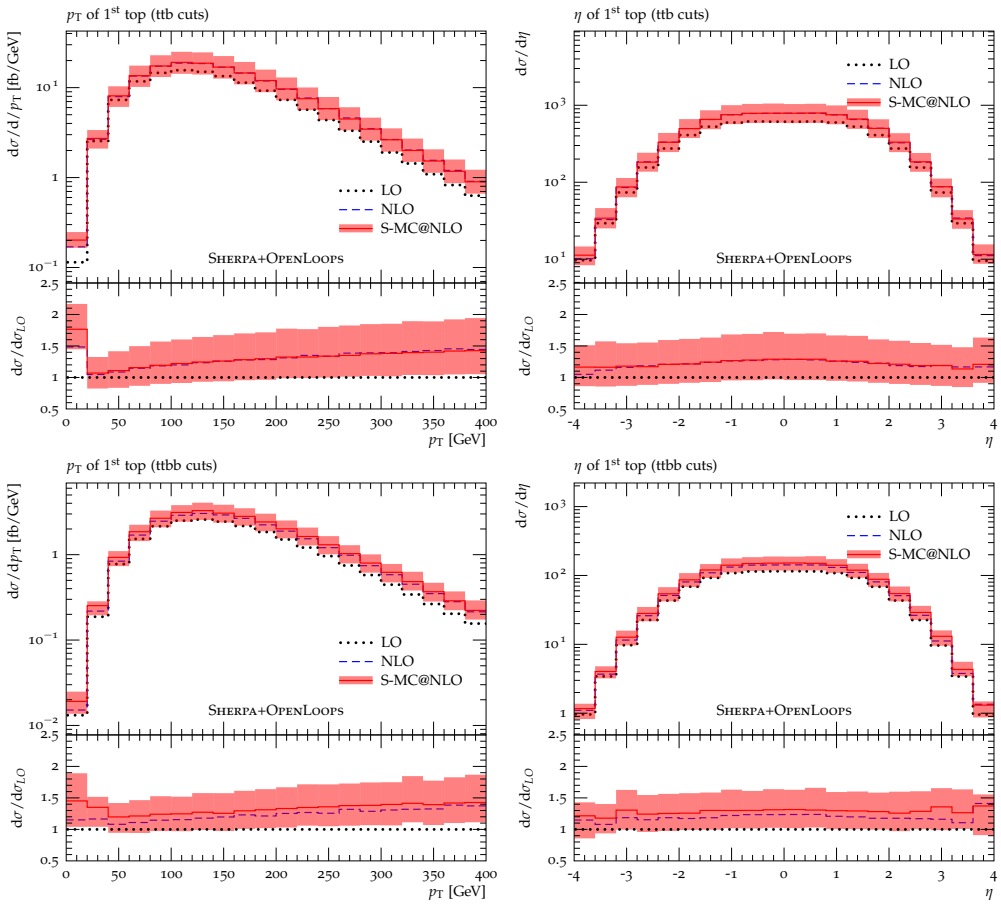


Figure 2.4: Transverse momentum and rapidity distribution of the hardest top in the $t\bar{t}b$ (top panel) and $t\bar{t}b\bar{b}$ (bottom panel) sample. The S-MC@NLO bands display the combination in quadrature of μ_R, μ_F and μ_Q scale variations.

As pointed out in [37], S-MC@NLO predictions exhibit a significant excess with respect to fixed-order NLO ones in b-jets correlation observables. This contribution comes from double $g \to b\bar{b}$ collinear splittings, where one splitting is produced by matrix elements with NLO accuracy and the second one takes place at the level of the parton shower. The enhancement grows at high m_{bb} invariant mass because in that region the production of two back-to-back gluons, which then split into two collinear $b\bar{b}$ pairs, is increasingly favoured with respect to the direct production of two hard b-quarks. This interpretation is supported by the fact that the S-MC@NLO excess increases

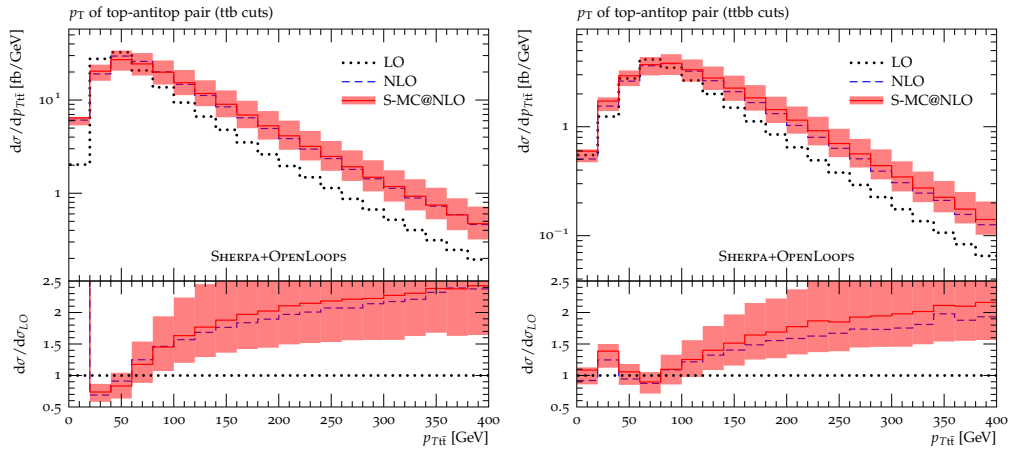


Figure 2.5: Transverse momentum distribution of the $t\bar{t}$ system in the $t\bar{t}b$ (left plot) and $t\bar{t}b\bar{b}$ (right plot) sample. The S-MC@NLO bands display the combination in quadrature of μ_R, μ_F and μ_Q scale variations.

at high di-jet invariant mass at a similar rate as the ratio of the $t\bar{t}gg$ to $t\bar{t}b\bar{b}$ cross sections. For instance, using LO matrix elements, we checked that both quantities increase by a factor two in the range 100 and 250 GeV (see Fig. 2.7). As discussed in [37], in order to exclude the possibility that the S-MC@NLO excess observed in our simulations is artificially enhanced by a too high choice of the resummation scale, we checked that the characteristic "double-splitting" effect in the $m_{b\bar{b}}$ distribution of Fig. 2.2 is present also in simulations based on merged LO matrix elements for $t\bar{t}$ plus multi-jet production. In this framework, $t\bar{t}b\bar{b}$ events are not showered with a global resummation scale, but starting from a scale that is determined according to the most likely shower history of the event at hand. Comparing the shape of the S-MC@NLO distribution of Fig. 2.8 against MEPS@LO simulations [26] of $t\bar{t}+ \leq 3j$ with massive b-quarks, we found good agreement for relatively low merging scales around 15 GeV, i.e. when the region associated to soft-collinear emissions is mostly described in terms of matrix elements, as in the S-MC@NLO simulation.

2.2.2 Additional sources of theoretical uncertainty

In the following we will extend the results of [37] focusing on theoretical uncertainties. Typically theoretical uncertainties associated with missing higher orders are assessed by factor-two variations of the renormalization and factorization scales around the central value. This variation quite often

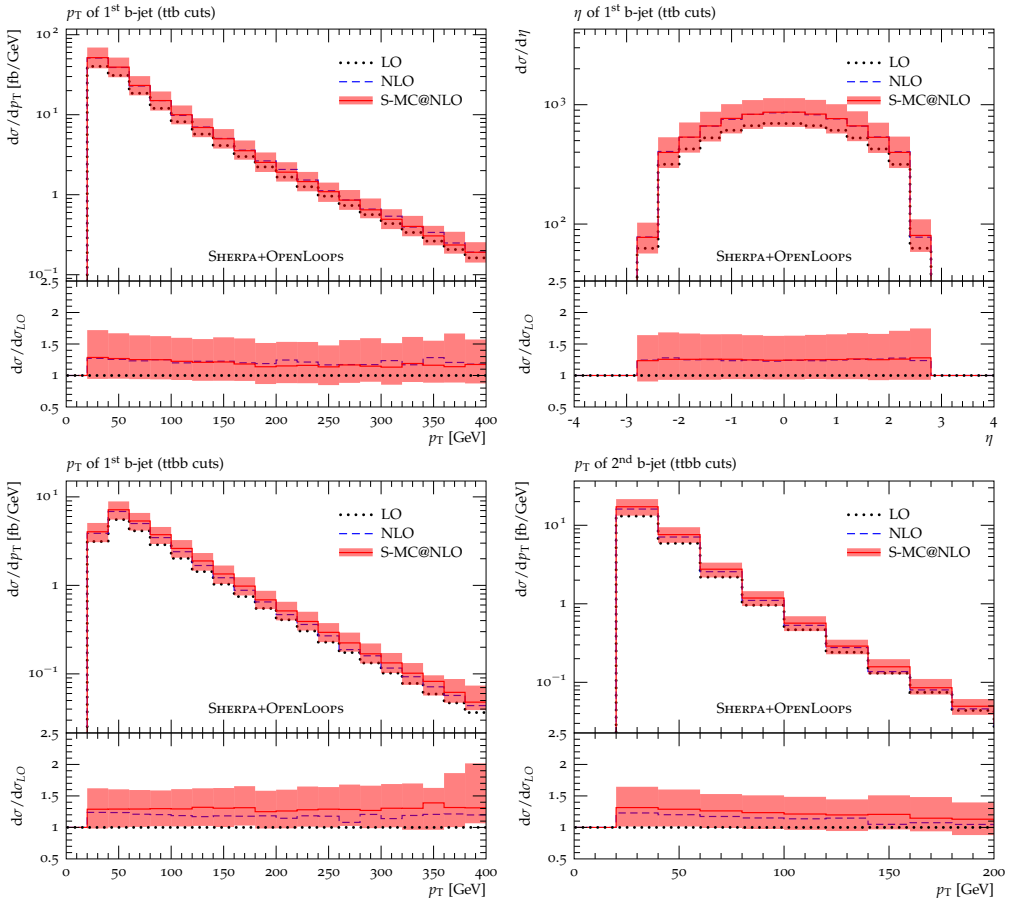


Figure 2.6: Transverse momentum and rapidity distribution of the hardest b-jet in the $t\bar{t}b$ sample (top panel) and transverse momentum distribution of the 1st and 2nd b-jets in the $t\bar{t}b\bar{b}$ sample (bottom panel). The S-MC@NLO bands display the combination in quadrature of μ_R, μ_F and μ_Q scale variations.

affects the normalisation of the observables, inducing only mild effects onto the shapes. The process at hand, $t\bar{t}b\bar{b}$ production, depends on the strong couplings constant as α_S^4 at leading order and is characterised by an order 20%–30% theoretical uncertainty at next-to-leading order (see Table 2.1). However, the multi-parton and multi-scale nature of the process entails a very wide freedom in the choice of the scales, that in principle span almost two orders of magnitudes (from m_b to $m_{t\bar{t}}$). In the context of experimental analyses, data driven approaches tend to "fit out" normalisation uncertainties, lowering de facto the impact of theoretical uncertainties assessed by the typical factor-two variations, which can then result in an underestimate of

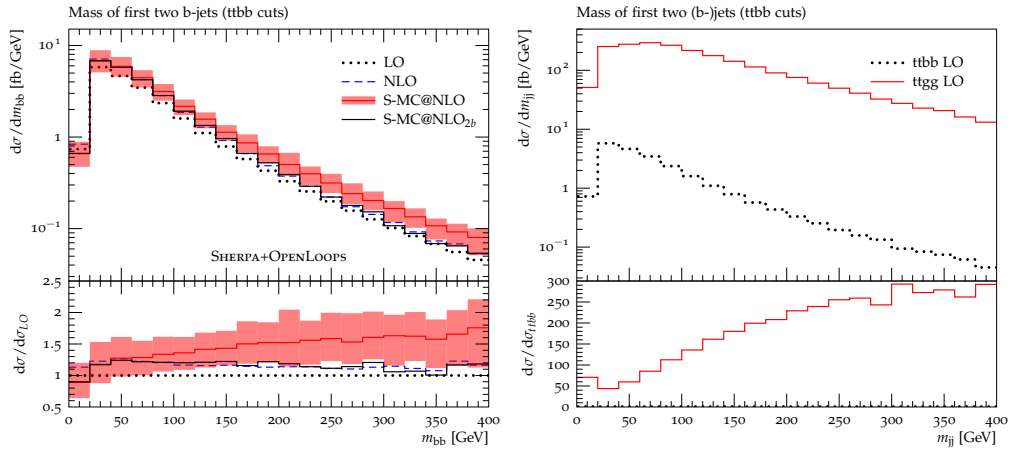


Figure 2.7: Invariant mass distribution of the first two b-jets in the $t\bar{t}b\bar{b}$ sample (left plot), and di-jet invariant mass distributions for $pp \to t\bar{t}b\bar{b}$ and $pp \to t\bar{t}gg$ based on fixed-order LO matrix elements (right plot). The S-MC@NLO$_{2b}$ curve is obtained by switching off g $\to b\bar{b}$ splittings in the parton shower.

associated uncertainties, unless theoretical shape uncertainties are properly taken into account. Therefore, in this section, we perform an additional study of the scales dependence of the process, focusing on shape distortions induced by different scale and PDFs choices.

To this end we adopt the setup of the previous section as reference, and we consider the shape distortions caused by several independent variations. Concerning scale variations, we consider kinematic distortions of μ_R, μ_F and μ_Q using various combinations of the following variables

$$\mu_{CMMPS} \equiv \sqrt[4]{\prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}}, \qquad m_{b\bar{b}},$$
$$H_{T,b(t)} \equiv E_{T,b(t)} + E_{T,\bar{b}(\bar{t})}, \qquad H_T \equiv H_{T,b} + H_{T,t}.$$

Specifically we consider three classes of variations: global (*glo*) variations, where we set the same scale for μ_R, μ_F and μ_Q, renormalisation and resummation (resp. R and Q) scale variations. The performed variations and the respective cross sections are listed in Table 2.2.

The dominant effects are driven by μ_R variation. Moving from a soft (μ_{CMMPS}) to an hard (m_t or H_T) scale yields a 30–40% effect on the normalisation, which is close to the NLO theoretical uncertainty based on standard factor-two rescalings. Some variations, despite giving milder effects, affect

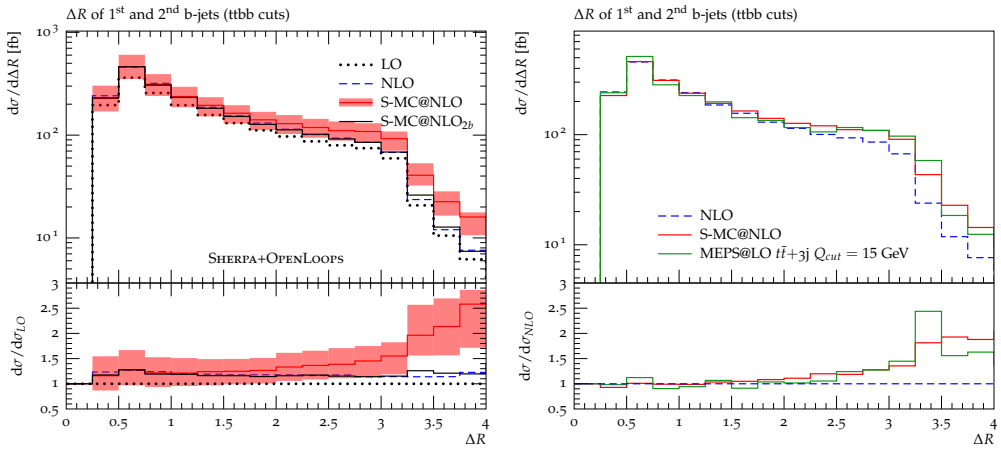


Figure 2.8: ΔR distribution of the first two b-jets in [37] (left plot), and with a MEPS@LO $t\bar{t}+\leq 3j$ simulation in the 4F scheme with $Q_{cut} = 15\,\text{GeV}$ (right plot) in the $t\bar{t}b\bar{b}$ sample. The S-MC@NLO$_{2b}$ curve is obtained by switching off $g \to b\bar{b}$ splittings in the parton shower.

Scale	**default**	glo-HT	glo-Mt	glo-soft	R-Mbb	R-HTb	R-HTt	Q-CMMPS	Q-Mt
μ_R	μ_{CMMPS}	$H_T/2$	m_t	μ_{CMMPS}	$\sqrt{m_t m_{b\bar{b}}}$	$\sqrt{m_t H_{T,b}/2}$	$\sqrt{m_t H_{T,t}/2}$	μ_{CMMPS}	μ_{CMMPS}
μ_F	$H_{T,t}/2$	$H_T/2$	m_t	μ_{CMMPS}	$H_{T,t}/2$	$H_{T,t}/2$	$H_{T,t}/2$	$H_{T,t}/2$	$H_{T,t}/2$
μ_Q	$H_{T,t}/2$	$H_T/2$	m_t	μ_{CMMPS}	$H_{T,t}/2$	$H_{T,t}/2$	$H_{T,t}/2$	μ_{CMMPS}	m_t
Cuts	σ_{ref} [pb]	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$	$\Delta\sigma/\sigma_{ref}$
ttb	3313	−41%	−27%	+4.7%	+2.3%	+1.1%	−32%	−3.5%	−0.3%
$ttbb$	600	−33%	−17%	−0.7%	+0.2%	+3.4%	−22%	−6.4%	−1.1%
$ttbb_{100}$	181.0	−29%	−13%	−9.2%	−5.6%	+2.5%	−17%	−14%	−2.9%

Table 2.2: List of the performed variations and related cross sections. The results are shown as the relative deviations with respect to the default setup.

differently ttb, $ttbb$ and $ttbb_{100}$ rates, inducing a sort of shape variation. This is the case for example for the glo-soft and Q-CMMPS variations, where this kind of effect is 10–15%.

Another class of variation concerns PDFs and bottom mass (m_b) variations. PDFs shape effects are assessed by employing a different PDF set

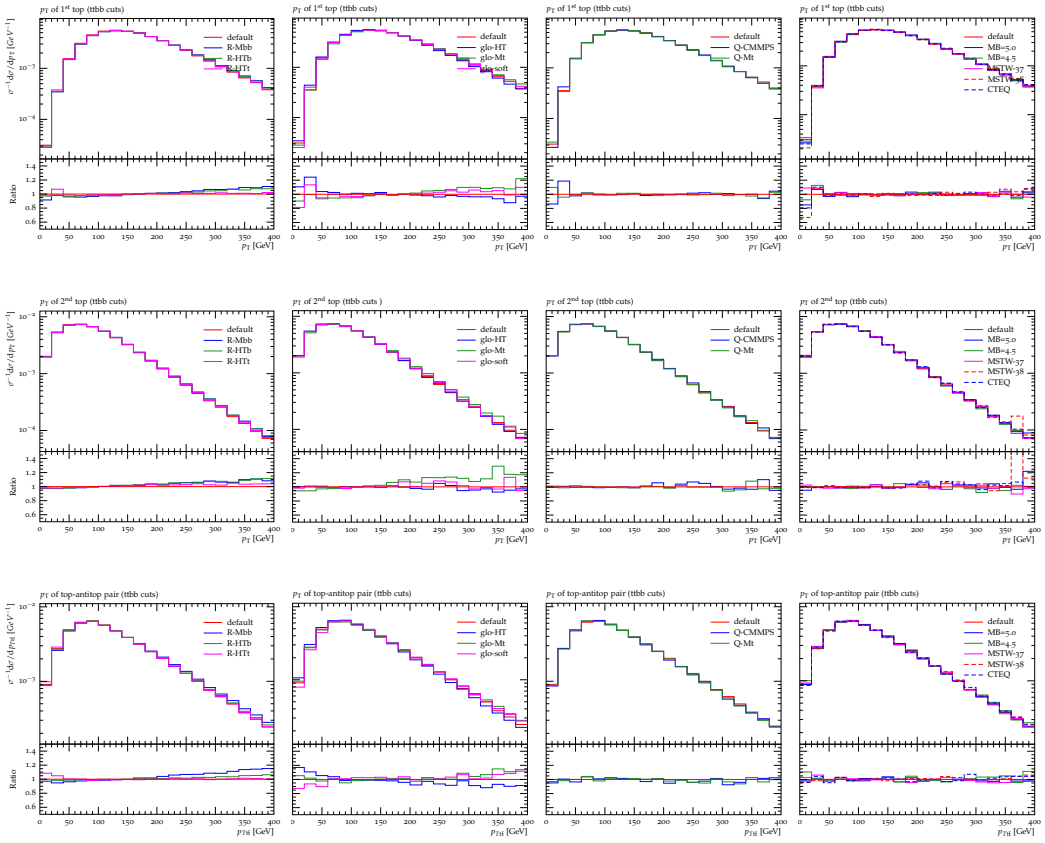


Figure 2.9: Shape uncertainties of transverse momentum of the hardest top (top panel), the softest top (middle panel) and the $t\bar{t}$ system (bottom panel) with $t\bar{t}b\bar{b}$ cut.

(CTEQ10nlo [120]) and different PDFs members of the default set [6]. Despite PDF uncertainties are not expected to affect significantly shapes, bottom mass variations are potentially more relevant, especially in configurations where two bottoms become collinear.

The shape distortions induced by all considered classes of variations are shown in Figs.2.9, 2.10 and 2.11, for various observables. The overall effect of the different variations amounts to a 10% shape distortion on the top p_T distribution (Fig 2.9), which grows to 20% in the tails, mainly due to the dependence of the renormalisation scale on the top kinematics. In Fig 2.10 the effects on b-jet transverse momentum are shown. Here the distortion

[6] In particular, we focused on the MRSTW2008 37 and 38 PDFs members, since these embody the maximum variation of the gluon distribution.

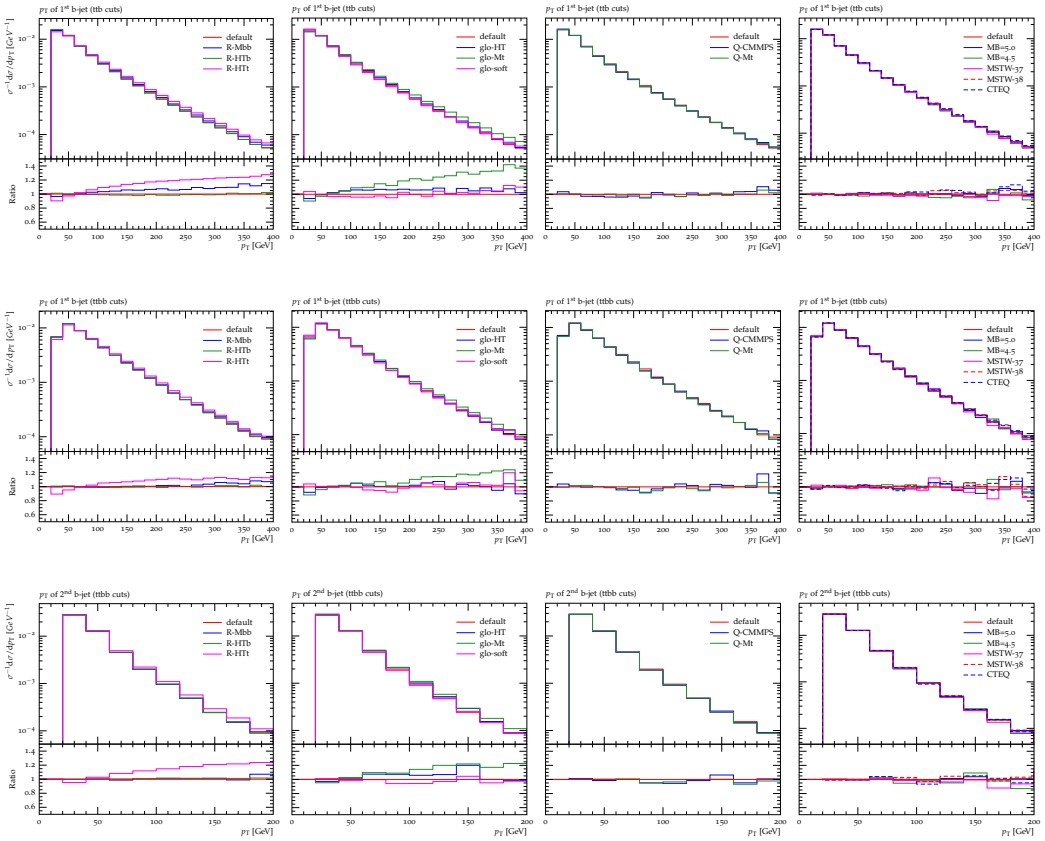


Figure 2.10: Shape uncertainties of transverse momentum of the 1st bjet with $t\bar{t}b$ cut (top panel), the same with $t\bar{t}b\bar{b}$ cut (middle panel) and of the 2nd bjet with $t\bar{t}b\bar{b}$ cut (bottom panel).

is $\sim$ 10–20% which increases up to 40% in the tails, driven by the b–jet dependence of μ_R. Finally, in the b–jet correlations, which are very important for $t\bar{t}H(b\bar{b})$ analyses, there is a $\sim$ 10–20% effect in the collinear region (small ΔR and m_{bb}), and a 10–20% reduction in m_{bb} when a softer resummation scale is employed (Fig. 2.11).

All predictions have been obtained with the SHERPA+OPENLOOPS framework and are publicly available for experimental analyses as Monte Carlo samples [121]. These event samples include top decays, multi-parton interactions, hadronisation and hadron decays, and thus they are ready-to-use for data analysis. They have already been employed in the $t\bar{t}H(b\bar{b})$ analysis by ATLAS in [13].

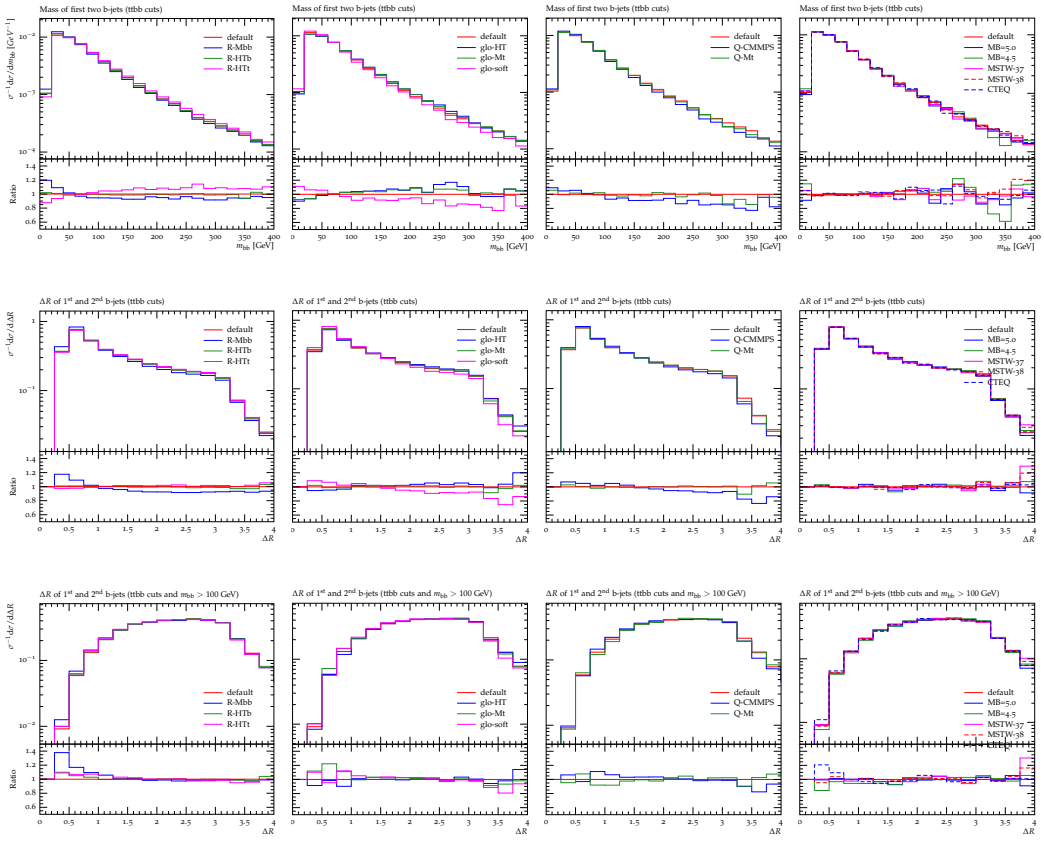


Figure 2.11: Shape uncertainties of the invariant mass (top panel) and on the ΔR separation (middle panel) of the first two b–jets with $t\bar{t}b\bar{b}$ cut, and of the ΔR separation with $t\bar{t}b\bar{b}_{100}$ cut (bottom panel).

2.3 Resummation scale dependence

The matching methods discussed in section 1.3 guarantee that QCD radiation arising from the parton shower does not overlap with real emission at NLO. In the POWHEG method, this is achieved through exponentiation of the real emission, while in the MC@NLO framework the first shower emission is matched to NLO matrix elements via appropriate subtraction terms. Both methods implement a splitting of the radiation phase space into a hard (non resummed) region and a soft (resummed) region. In the MC@NLO method and its S-MC@NLO variant, this splitting is defined through the resummation scale μ_Q, which also corresponds to the parton shower starting scale. In this section we will address the dependence of MC@NLO simulations of $t\bar{t}b\bar{b}$ production with respect to μ_Q.

In the framework of an extensive comparison of $t\bar{t}b\bar{b}$ simulations based on different NLO+PS matching methods and tools [122] it was pointed out that the choice of μ_Q and variations thereof might lead to sizeable uncertainties in the MC@NLO approach. The abovementioned study [122] is presented in Appendix B and will appear in the fourth Yellow Report of the LHC Higgs Cross Section Working Group [122] (denoted as YR4 in the following). The study of [122] presents a tuned comparison of NLO+PS simulations of $t\bar{t}b\bar{b}$ production at 13 TeV performed with SHERPA+OPENLOOPS, MG5aMC@NLO+PHYTIA 8 (denoted as MG5+PHYTIA 8 in the following) and POWHEL + PHYTIA 8. In this context, as discussed in detail in Appendix B, it was observed that SHERPA+OPENLOOPS and POWHEL + PHYTIA 8 predictions, which rely on different flavour number schemes, different matching methods, and different parton showers, are well in agreement for a wide range of observables. In contrast, the comparison of SHERPA+OPENLOOPS against MG5+PHYTIA 8, both based on the four-flavour number scheme and the MC@NLO matching method, has revealed sizeable discrepancies in observables that are sensitive to QCD radiation at NLO. In particular, it turned out that such discrepancies are especially sensitive to the choice of μ_Q in MG5+PHYTIA 8.

Before we address this issue in detail, let us recall the scale choices that have been adopted for the simulation of $pp \to t\bar{t}b\bar{b}$ in the YR4 [122]. For all other aspects of the setup we refer to Appendix B. The factorisation scale μ_F and the renormalisation scale μ_R have been set to

$$\mu_R = \left(\prod_{i=t,\bar{t},b,\bar{b}} E_{T,i} \right)^{\frac{1}{4}} \qquad \mu_F = \frac{H_T}{2} = \frac{1}{2} \sum_{i=t,\bar{t},b,\bar{b},j} E_{T,i}, \tag{2.6}$$

where $E_{T,i} \equiv \sqrt{M_i^2 + p_{T,i}^2}$ denotes the top and bottom transverse mass defined at parton level. Theoretical uncertainties have been assessed via standard variations $\xi_R\mu_R$ and $\xi_F\mu_F$ with $1/2 < \xi_R, \xi_F < 2$.

Since initial-state evolution as implemented in the parton shower amounts to backward PDF evolution, in the MC@NLO framework it is natural to identify μ_Q with μ_F. Thus the SHERPA+OPENLOOPS simulation was performed using $\mu_Q = \mu_F = \frac{H_T}{2}$. However, in the MG5+PHYTIA 8 [27] framework only resummation scales of the form $\mu_Q = \xi\sqrt{\hat{s}}$ can be chosen, where the prefactor ξ is randomly distributed in the freely adjustable range $[\xi_{min}, \xi_{max}]$ (set to $[0.1, 1]$ at default), with a distribution that is strongly peaked at $(\xi_{min} + \xi_{max})/2$. In order to achieve a reasonably consistent resummation scale setting in the comparison of SHERPA+OPENLOOPS and MG5+PHYTIA 8, it was decided to lower the ξ upper bound in the latter from its default value

of one to $\xi_{max} = 0.25$. This was motivated by the fact that the resulting μ_Q expectation value is reasonably close to the one of $H_T/2$. The choice $\xi_{max} = 0.25$ for processes with b-jets in the final states is also supported by the analysis of $b\bar{b}H$ production in [123].

Given the sizeable differences between SHERPA+OPENLOOPS and MG5+PHYTIA 8 predictions in [122] and the fact that such differences are even much more pronounced when using the default setting $\xi_{max} = 1$ in MG5+PHYTIA 8, it is important to address the sensitivity of MC@NLO predictions, both in MG5+PHYTIA 8 and SHERPA+OPENLOOPS, with respect to variations of the resummation scale. To this end, in the following, we complement the study of [122] with additional SHERPA+OPENLOOPS and MG5+PHYTIA 8 results based on a μ_Q choice that corresponds to the default in MG5+PHYTIA 8. This is achieved by setting $\mu_Q = 0.5\sqrt{\hat{s}}$ in SHERPA+OPENLOOPS[7]. Predictions corresponding to the resummation scale choice adopted in the YR4 study [122] will be labelled as SHERPA+OPENLOOPS ($\mu_Q = H_T/2$) and MG5+PHYTIA 8 ($\xi_{max} = 0.25$), while their resummation scale variations will be labelled as $\texttt{S+OL}_{\sqrt{\hat{s}}/2}$ ($\mu_Q = \sqrt{\hat{s}}/2$) and $\texttt{MG5}_{\sqrt{\hat{s}}/2}$ ($\xi_{max} = 1$). Comparing these different predictions provides new insights into the μ_Q dependence of the individual predictions and in the discrepancy between the different tools.

Results for the inclusive b-jet multiplicity distribution in $t\bar{t} + b$-jet production at 13 TeV are presented in Table 2.3 and Fig. 2.12. Specifically we consider b-jets with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ and report $t\bar{t} + b$-jets cross sections with $n_b \geq N_b$ b-jets for various values of N_b. In the following we will refer to the results for $N_b = 1, 2, 3, 4$ as ttb, $ttbb$, $tt + 3b$ and $tt + 4b$ cross sections, respectively. At the level of the inclusive b-jet multiplicity distribution, the increase of the resummation scale affects the simulations differently: $\texttt{S+OL}_{\sqrt{\hat{s}}/2}$ exhibits very good stability, showing an excess w.r.t. the nominal YR4 simulation of only +2% for $N_b = 2$ and up to +22% for $N_b = 4$. Conversely, in $\texttt{MG5}_{\sqrt{\hat{s}}/2}$ the enhancement resulting from the change in μ_Q amounts to +10% for $N_b = 2$ and about +40% for $N_b = 3$. In this respect one should note that the cross sections with more than 3 and 4 b-jets get contributions from the additional b–quarks produced by the parton shower, and are thus potentially much more sensitive to μ_Q with respect the ones with only two or less b-jets.

In differential distributions, we observe significant μ_Q sensitivity already with ttb cuts (Fig. 2.13). In this case, the S-MC@NLO implementation in SHERPA turns out to be stable w.r.t. the increase of the resummation scale, and top-quark and b-jet observables show only effects of order +10% or less

[7]Strictly speaking, the μ_Q distribution with default choices in MG5+PHYTIA 8 is strongly peaked at $(0.1 + 1)\sqrt{\hat{s}}/2 = 0.55\sqrt{\hat{s}}$, but the difference is almost negligible.

in most cases. Slightly larger effects arise in observables involving the first light jet emission. The first light jet p_T shows an enhancement of 20–25% in the tail w.r.t. corresponding predictions with nominal μ_Q. This is well consistent with the level of uncertainty expected from conventional factor-two scale variations. The μ_Q dependence in $\texttt{MG5}_{\sqrt{\hat{s}}/2}$ turns out to be much more pronounced. It amounts to +40% in the high p_T tail and reaches more than +100% at $p_T \sim 200$GeV. Looking at the ΔR separation between the light jet and the hardest b–jet, the enhancement at $R \sim \pi$ suggests that the additional radiation is emitted in the opposite direction of the hardest $b-$jet.

Differential distributions subject to *ttbb* cuts, shown in Fig. 2.14, feature even larger μ_Q sensitivity. Again, $\texttt{S+OL}_{\sqrt{\hat{s}}/2}$ results show only small deviations w.r.t. their nominal counterpart. Top-quark and b-jet observables exhibit excesses lower than 20% almost everywhere, and it is interesting to observe that increasing the resummation scale has a tiny impact on the invariant mass of the b–jets pair, which is a crucial observable in Higgs searches. In the tail of the p_T spectrum of the hardest light-jet the μ_Q variation reaches up to 40%, but for this and all other observables the effects are well below renormalisation and factorisation scale variations. In the case of the $\texttt{MG5}_{\sqrt{\hat{s}}/2}$ simulation, we observe moderate μ_Q variations ($\sim +10\%$) in the top-quark observables, but quite significant enhancements in $b-$jet distributions. Nonetheless, the invariant mass of the b–jets pair exhibits a mild excess (10–20%), well consistent with standard scale variations. In contrast, very large effects are observed in the hardest light-jet spectrum, where the μ_Q variation reaches +100% at $p_T \sim 200$GeV and even more in the tail.

In summary, the S-MC@NLO method implemented in SHERPA turns out to be very stable w.r.t. the choice of the resummation scale, with effects of 10–20% or less on most of the observables that are interesting in the context of Higgs searches. A bigger effect arises in distributions which are sensitive to the first shower emission, such as the p_T spectrum of the hardest light jet, where the deviation reaches +40% in the hard tail, which is anyway within standard scale variations. Despite the formal equivalence of the two methods at order α_S, the MC@NLO method implemented in MG5aMC@NLO exhibits a stronger shower dependence in a wide class of observables. Increasing μ_Q enhances the shower activity, and the resulting effects shows up already in rather inclusive observables, such as the transverse momentum of the hardest $b-$jet with *ttb* cuts. The origin of this effect lies in the enhancement of the first jet emission, by up to +100%, in the intermediate p_T range. These findings suggest that the sizeable discrepancies reported in the study of [122] (see Appendix B) might be related to the strong μ_Q sensitivity of the MC@NLO implementation in MG5aMC@NLO+PHYTIA 8 rather than to intrinsic features of the MC@NLO approach.

Simulation	$n_b \geq 1$[fb]	$n_b \geq 2$ [fb]	$n_b \geq 3$ [fb]	$n_b \geq 4$ [fb]
SHERPA+OPENLOOPS	$12939^{+30\%}_{-27\%}$	$2413^{+21\%}_{-24\%}$	$71.3^{+34\%}_{-28\%}$	$6.3^{+19\%}_{-23\%}$
S+OI$_{\sqrt{\hat{s}}/2}$	13073	2477	82.4	7.7
$\sigma_{\sqrt{\hat{s}}/2}/\sigma_{\text{nominal}}$	1.01	1.03	1.16	1.22
MG5aMC@NLO+PHYTIA 8	$13833^{+37\%}_{-29\%}$	$3192^{+38\%}_{-29\%}$	$116.4^{+33\%}_{-28\%}$	$20.8^{+42\%}_{-31\%}$
MG5$_{\sqrt{\hat{s}}/2}$	14224	3400	166.5	24.5
$\sigma_{\sqrt{\hat{s}}/2}/\sigma_{\text{nominal}}$	1.02	1.07	1.43	1.18

Table 2.3: SHERPA+OPENLOOPS and MG5aMC@NLO+PHYTIA 8 predictions at NLO+PS accuracy for integrated $t\bar{t} + b-$jets cross sections at 13 TeV in bins with $n_b \geq 1, 2, 3, 4$, and the effect of higher resummation scale. Predictions obtained with nominal μ_Q choice and $\mu_Q = \sqrt{s}/2$ are compared. In the nominal predictions also the theoretical uncertainty is shown, obtained with standard μ_R and μ_F variations.

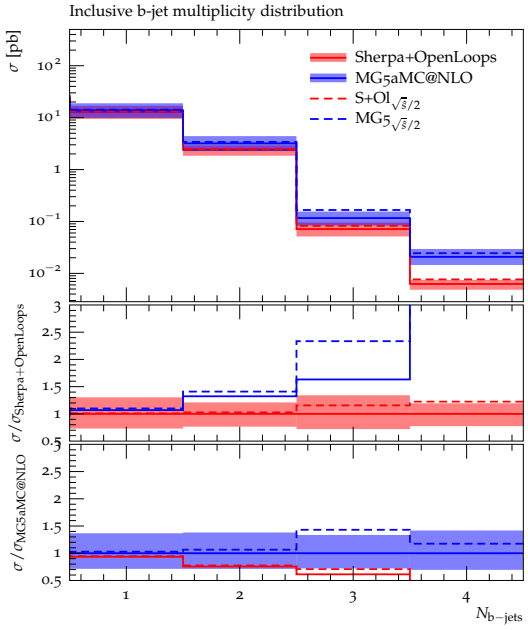


Figure 2.12: NLO+PS predictions for integrated $t\bar{t} + b-$jets cross sections at 13 TeV in inclusive bins with $n_b \geq N_b$ b jets. Predictions obtained with nominal μ_Q choice and $\mu_Q = \sqrt{s}/2$ are compared. The ratio plots show the effect of μ_Q variation and standard scale variations normalised to nominal predictions in SHERPA+OPENLOOPS and MG5aMC@NLO+PHYTIA 8.

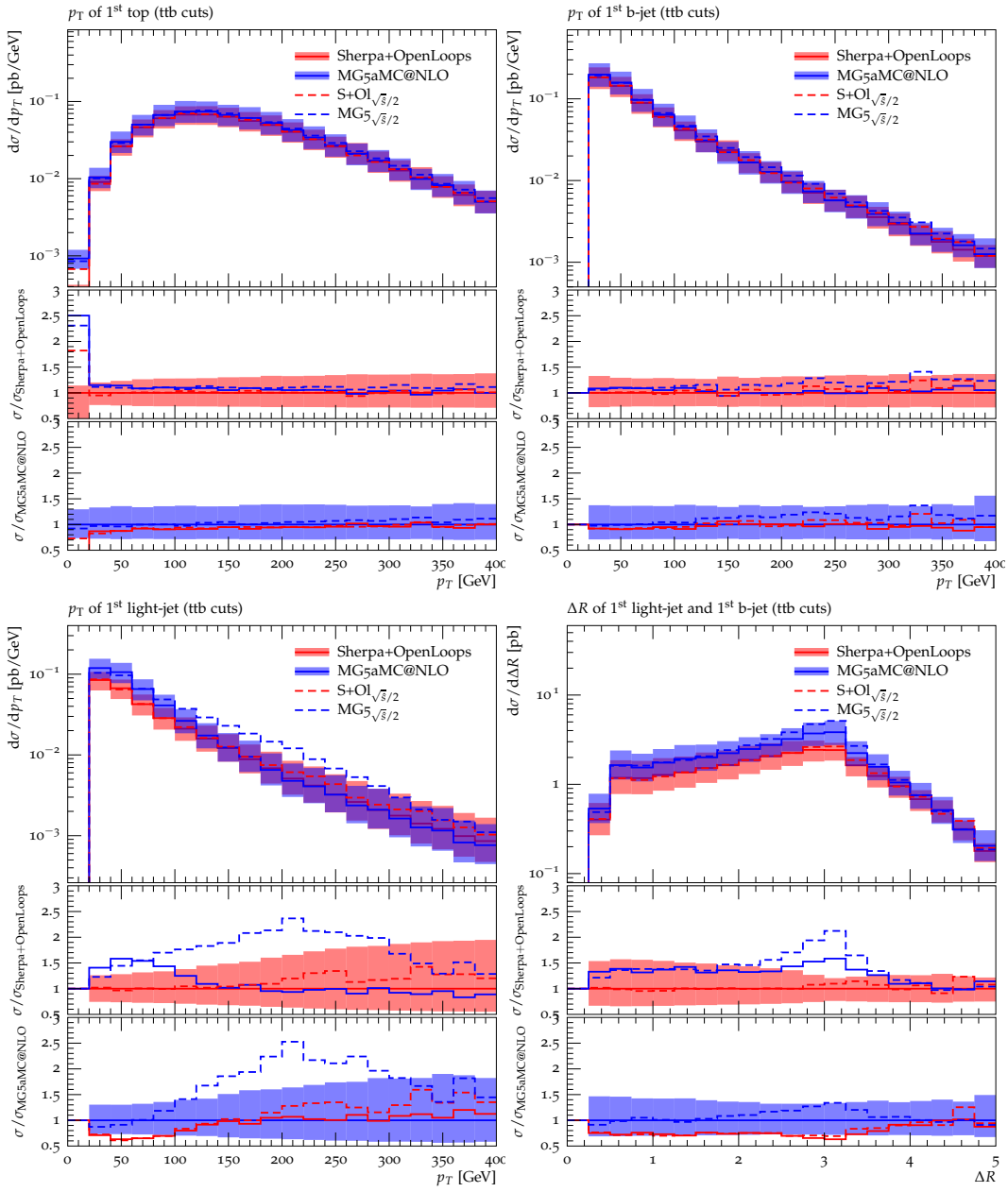


Figure 2.13: Distributions for $pp \to t\bar{t}+ \geq 1b$ jets at 13 TeV. Ratio plots like in Fig. 2.12.

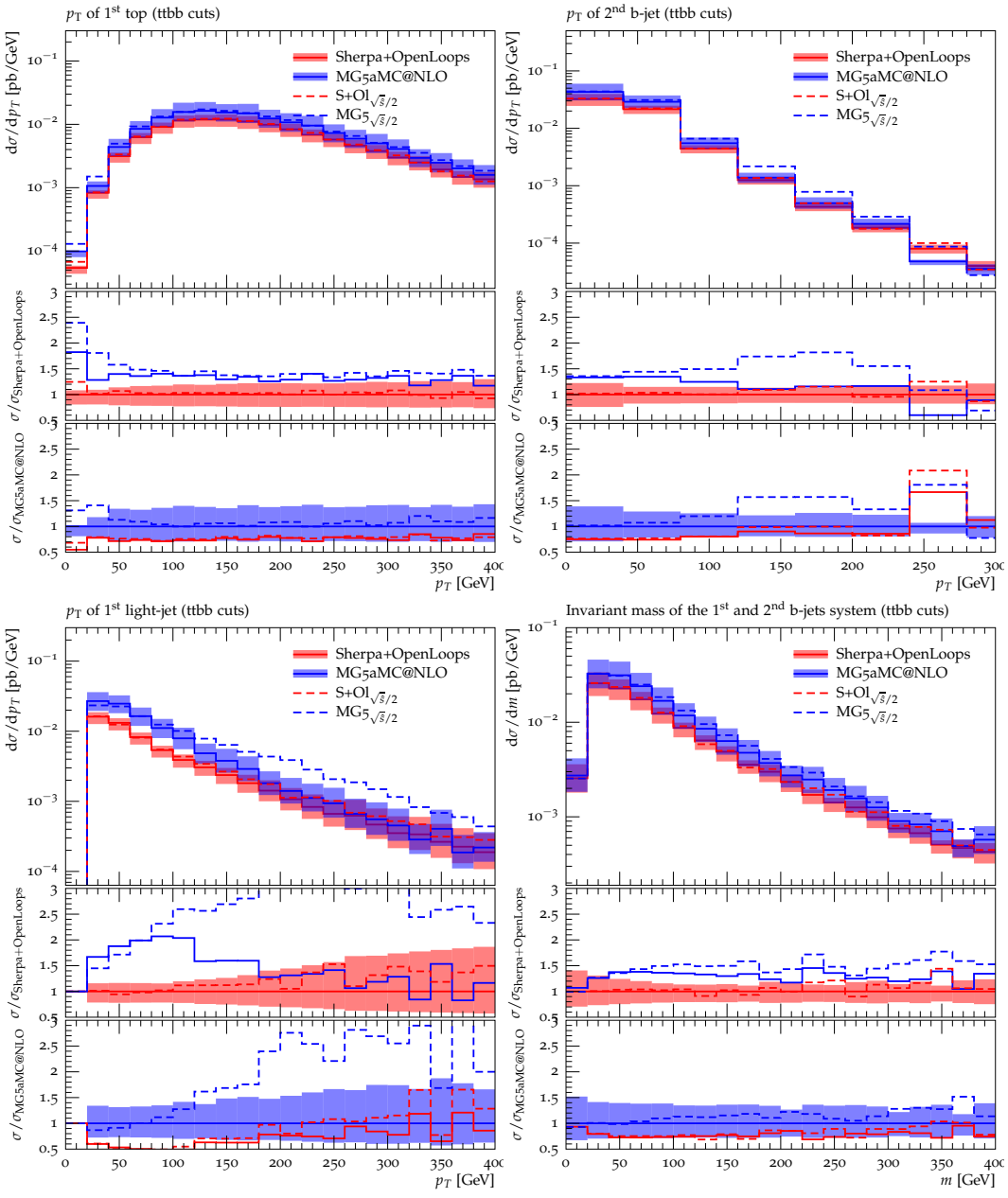


Figure 2.14: Distributions for $pp \to t\bar{t}+ \geq 2b$ jets at 13 TeV. Ratio plots like in Fig. 2.12.

Chapter 3

Next-to-leading order QCD predictions for top-quark pair production with up to three jets

3.1 Introduction

The top quark as the heaviest known elementary particle plays a fundamental role, both in the Standard Model and in new physics scenarios. Experimental analyses of Large Hadron Collider (LHC) data collected during run II will provide unprecedented reach at high energy and in exclusive phase space regions with associated production of jets and vector bosons or Higgs bosons. The production of a $t\bar{t}$ system in association with multiple jets plays an especially important role as a background to new physics searches and to various Higgs and Standard Model analyses. In particular, the precise theoretical control of $t\bar{t}$+multijet backgrounds is one of the most important prerequisites for the observation of top-quark production in association with a Higgs boson, which would give direct access to the top-quark Yukawa coupling. In addition, $t\bar{t}$+multijet production allows for powerful test of perturbative QCD and is also routinely exploited for the validation of Monte Carlo tools that are used in a multitude of LHC studies. All these analyses require theoretical predictions at the highest possible accuracy.

Inclusive top-quark pair production at hadron colliders has been computed fully differentially to next–to–next–to–leading order (NNLO) in the strong coupling expansion [124, 125]. Predictions for top-quark pair production in association with up to two jets are available at the next–to–leading

order (NLO) [29, 30, 33–35, 126], and NLO calculations for inclusive top-quark pair production and in association with up to one or two jets were matched to parton showers in order to provide predictions at the particle level [15, 24, 31, 32, 36, 37, 127–130].

In this letter we report on the first computation of top-quark pair production with up to three jets at NLO QCD. At present only few scattering processes with more than six external legs are known at NLO [131–138], and the calculation at hand is the first one that deals with a $2 \to 5$ process with seven colored external particles including also heavy quarks. Detailed predictions are presented for $pp \to t\bar{t} + 0, 1, 2, 3$ jets with stable top quarks at 13 TeV, both at the level of cross sections and differential distributions. We also investigate the scaling behavior of $t\bar{t}$+multijet cross sections with varying jet multiplicity.

The characteristic scales of $t\bar{t}$+multijet production, i.e. the invariant mass of the $t\bar{t}$ system and the transverse momentum threshold for jet production, are typically separated by more than one order of magnitude, while differential observables involve multiple scales, which can be distributed over more than two orders of magnitude. In this situation, finding renormalisation and factorisation scales that ensure a decent convergence of perturbative QCD for the widest possible range of observables is not trivial. Moreover, in the presence of a wide spectrum of scales, the usage of standard factor-two variations for the estimation of theoretical uncertainties due to missing higher-order effects becomes questionable. Motivated by these observations, to gain more insights into the scale dependence of $t\bar{t}$+multijet production and related uncertainties we compare a fixed-order calculation, with the standard scale choice $H_{\mathrm{T}}/2$, against results based on the MiNLO method [38]. The scale $H_{\mathrm{T}}/2$ was found to yield stable and reliable NLO predictions for V+multijet production [139], while the MiNLO method is especially well suited for multi-scale QCD processes, as it controls, through next-to-leading logarithmic[1] (NLL) resummation, the various higher-order logarithms that emerge from soft and collinear effects in the presence of widely separated scales. The present study provides a first systematic comparison of the two approaches.

[1] More precisely, the MINLO approach guarantees a consistent resummation of all NLL effects apart from colour-suppressed contributions stemming from soft wide angle radiation, which enter $pp \to t\bar{t}$+jets as well as any processes with more than three coloured external partons.

partonic channel \ N	0	1	2	3
$gg \to t\bar{t} + N\,g$	47	630	9'438	152'070
$u\bar{u} \to t\bar{t} + N\,g$	12	122	1'608	23'835
$u\bar{u} \to t\bar{t}u\bar{u} + (N-2)\,g$	–	–	506	6'642
$u\bar{u} \to t\bar{t}d\bar{d} + (N-2)\,g$	–	–	252	3'321

Table 3.1: Number of one-loop Feynman diagrams in representative partonic channels in $pp \to t\bar{t} + N\,\text{jets}$ for $N = 0, 1, 2, 3$.

3.2 Details of the calculation

Our calculations are performed using the event generator SHERPA [28,140] in combination with OPENLOOPS [71,96], a fully automated one-loop generator based on a numerical recursion that allows the fast evaluation of scattering amplitudes with many external particles. For the reduction to scalar integrals and for the numerical evaluation of the latter we used CUTTOOLS [102] in combination with ONELOOP [103] and, alternatively, the COLLIER library [101], which implements the methods of [98–100]. Tree amplitudes are computed using COMIX [79], a matrix-element generator based on the color-dressed Berends-Giele recursive relations [84]. Infrared singularities are canceled using the dipole subtraction method [46, 115], as automated in COMIX, with the exception of K- and P-operators that are taken from the implementation described in [88]. COMIX is also used for the evaluation of all phase-space integrals. Analyses are performed with the help of RIVET [63].

We carry out a series of $pp \to t\bar{t} + N\,\text{jet}$ NLO calculations with $N = 0, 1, 2, 3$, taking into account the exact dependence on the number of colors, $N_c = 3$. As an illustration of the rapid growth of complexity at high jet multiplicity, in Table 3.1 we list the number of one-loop Feynman diagrams that contribute to a few representative partonic channels. In addition to the presence of more than 10^5 loop diagrams in the $gg \to t\bar{t} + 3g$ channel, we note that also the very large number of channels not listed in Table 3.1 as well as the computation of real contributions pose very serious challenges in the $t\bar{t} + 3\,\text{jet}$ calculation.

Proton–proton cross sections are obtained by using, both at LO and NLO, the CT14 NLO PDF set [64] with five active flavors, and the corresponding strong coupling. Matrix elements are computed with massless b-quarks, and top-quarks are kept stable. Hence, our results can be compared to data only upon reconstruction of the $t\bar{t}$ system and extrapolation of fiducial measurements to the full phase space. However, we expect the main features shown in our analysis to be present also in computations including top-quark decays

and acceptance cuts. The latter will undoubtedly play a role, but the reduction of scale uncertainties is generic as long as the radiative phase space is not heavily restricted by experimental cuts. Apart from performing a direct analysis, we also provide Root NTuples [141] that can be used in the future for more detailed studies including top-quark decays and matching to parton showers.

In our standard perturbative calculations we employ renormalisation and factorisation scales defined as $\mu_{\rm R} = \mu_{\rm F} = H_{\rm T}/2$, where $H_{\rm T} = \sum_i \sqrt{p_{{\rm T},i}^2 + m_i^2}$, with the sum running over all (anti)top quarks and light partons, including also real radiation at NLO. Results generated in this manner are compared to alternative computations based on the MiNLO procedure [38]. To this end, we have realized a fully automated implementation of the MiNLO method in Sherpa.

3.3 MiNLO method and implementation

The MiNLO method can be regarded as a generalized scale setting approach that guarantees a decent perturbative convergence for differential multi-jet cross sections. This is achieved via appropriate scale choices [142] and Sudakov form factors [143] that resum NLL enhancements in the soft and collinear regions of phase space. To this end, in the case of $t\bar{t}$+multijet production, LO partonic events of type $ab \to t\bar{t} + N$ partons are recursively clustered back to a core process $\tilde{a}\tilde{b} \to t\bar{t}$ by means of a $k_{\rm T}$ jet algorithm [58]. The resulting clustering history is interpreted as an event topology, where the N-jet final state emerges from the core process through a sequence of successive branchings that take place at the scales $q_N, \ldots, q_2, q_1$ and are connected by propagators. The nodal scales q_i correspond to the $k_{\rm T}$ measure of the jet algorithm, and only $1 \to 2$ branchings consistent with the QCD interaction vertices are allowed. In our implementation of the $k_{\rm T}$ jet algorithm we use the definition of ΔR given in Eq. (11) of [58] and we set $\Delta R = 0.4$. Typically, the $k_{\rm T}$ algorithm gives rise to ordered branching histories with $q_1 < \cdots < q_N < \mu_{\rm core}$, where $\mu_{\rm core}$ is the characteristic hard scale of the core process. However, also unordered branchings can occur. For instance, this can happen in the presence of jets with transverse momenta above $\mu_{\rm core}$. Since soft-collinear resummation does not make sense for such hard emissions, in our MiNLO implementation possible unordered clusterings are undone and alternative ordered configurations are considered. At the end, the branching history is restricted to ordered branchings $q_1 < \cdots < q_{\tilde{N}} < \mu_{\rm core}$, where $\tilde{N} = N - M$. The remaining M jets that can not be clustered in an ordered way are treated as part of the core process, and $\mu_{\rm core}$ is evaluated according

to the kinematics of the corresponding $t\bar{t} + M$ jet hard event.

At LO, the renormalisation scale $\mu_{\rm R}$ is chosen according to the event branching history in such a way that

$$[\alpha_{\rm s}(\mu_R)]^{N+2} = [\alpha_{\rm s}(\mu_{\rm core})]^{2+M} \prod_{i=1}^{\tilde{N}} \alpha_{\rm s}(q_i), \tag{3.1}$$

and in our calculation we set $\mu_{\rm core} = H_{\rm T}/2$.

The resummation of soft and collinear logarithms is achieved by dressing external and internal lines of the event topology by Sudakov form factors. At variance with the original formulation of MiNLO [38], in our implementation we employ the symmetry of the LO DGLAP splitting functions, $P_{ab}(z)$, to define physical Sudakov form factors

$$\begin{aligned}\Delta_a(Q_0, Q) = \exp\Bigg\{ - \int_{Q_0}^{Q} \frac{{\rm d}q}{q} \frac{\alpha_{\rm s}(q)}{\pi} \sum_{b=q,g} \\ \int_0^{1-q/Q} {\rm d}z \left(z\, P_{ab}(z) + \delta_{ab} \frac{\alpha_{\rm s}(q)}{2\pi} \frac{2C_a}{1-z} K \right) \Bigg\} \ ,\end{aligned} \tag{3.2}$$

where [51]

$$K = \left(\frac{67}{18} - \frac{\pi^2}{6} \right) C_A - \frac{10}{9} T_R\, n_f \ , \tag{3.3}$$

and $a = g, q$ corresponds to massless gluons and quarks, respectively. The representation (3.2) allows the interpretation of $\Delta_a(Q_0, Q)$ in terms of no-branching probabilities between the scales Q_0 and Q.

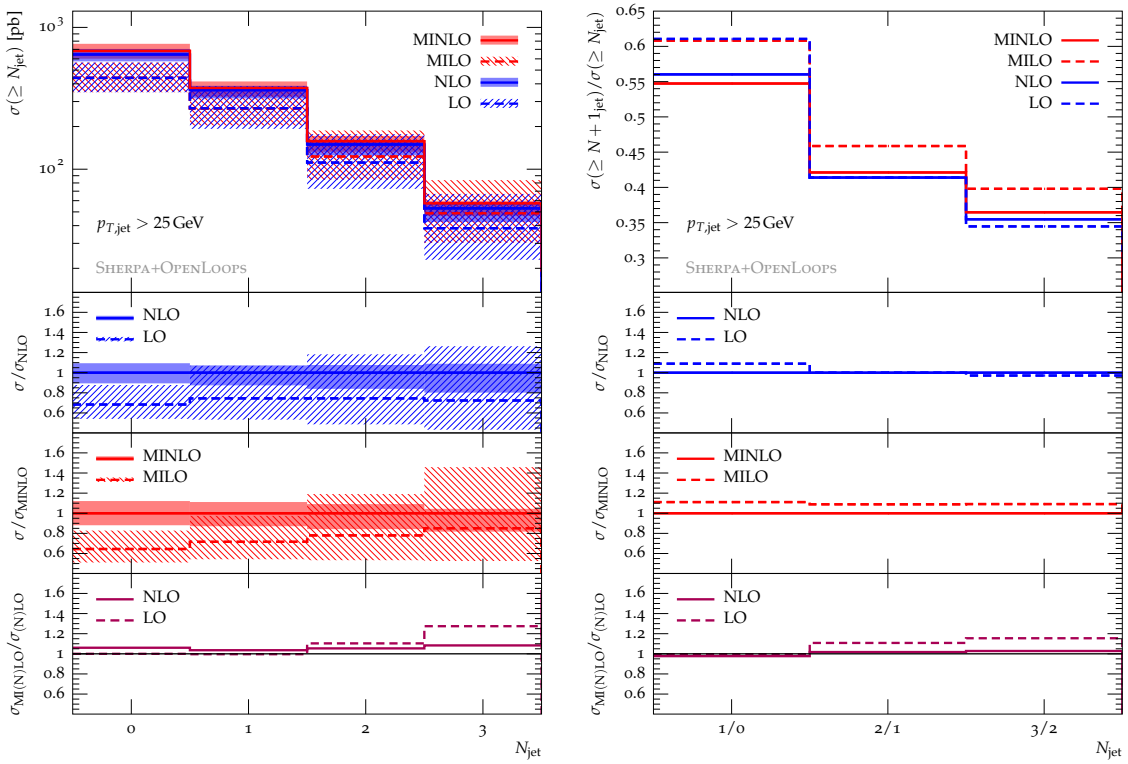
Given a LO event topology with $\tilde{N}$ ordered branchings, the lowest branching scale, $q_{\rm min} = q_1$, is identified as resolution scale, and the $\tilde{N}$ emissions are supplemented by Sudakov form factors that render them exclusive w.r.t. any extra emissions above $q_{\rm min}$. This is achieved by dressing each external line of flavor $a = q, g$ connected with the i-th branching by a form factor $\Delta_a(q_{\rm min}, q_i)$, while internal lines that connect successive branchings $k < l$ are dressed by factors $\Delta_a(q_{\rm min}, q_l)/\ \Delta_a(q_{\rm min}, q_k)$, which correspond to no-branching probabilities between q_k and q_l at resolution scale $q_{\rm min}$. For internal lines that connect branchings at q_k to the core process analogous no-branching probabilities between q_k and $\mu_{\rm core}$ are applied. Sudakov form factors along the incoming lines provide a NLL resummation that corresponds to the evolution of PDFs from the resolution scale $q_{\rm min}$ to the hard scale of the core process. Therefore, for consistency, PDFs are evaluated at the factorisation scale $\mu_{\rm F} = q_{\rm min}$.

The generalisation to NLO requires only two straightforward modifications of the LO algorithm. First, for what concerns the scale setting and Sudakov form factors, the contributions that live in the N-parton phase space,

i.e. Born and one-loop contributions as well as all IR-subtraction terms, are handled exactly as in LO. Instead, real-emission events that lead to histories with $\tilde{N}+1 \leq N+1$ ordered branchings at scales $q_0 < q_1 < \cdots < q_{\tilde{N}}$ are handled as Born-like $\tilde{N}$-parton events with resolution scale $q_{\text{min}} = q_1$, i.e. the softest branching at the scale q_0 is considered as unresolved and is simply excluded from the MiNLO procedure. In other words, the softest emission at NLO is not dressed with Sudakov form factors and does not enter the definitions of μ_{R} and μ_{F}. Second, appropriate counterterms are introduced in order to subtract the overall $\mathcal{O}(\alpha_{\text{s}})$ contribution from Sudakov form factors, such as to avoid double counting of NLO effects.

Concerning the treatment of top quarks a few extra comments are in order. Given the low rate at which top quarks radiate jets, such emissions are simply neglected in our implementation of the MiNLO procedure by excluding top quarks from the clustering algorithm. To quantify the uncertainty arising from this approach, we implemented an alternative algorithm that allows the combination of top quarks with other final-state partons in the massive Durham scheme [59, 60]. The difference between the two procedures is found to be about 10% at leading order and 5% at next-to-leading order for the observables studied here, and it is therefore smaller than the renormalisation and factorisation scale uncertainties. Finally, also the top quarks that enter the core process are dressed with Sudakov form factors $\Delta_t(q_{\text{min}}, \mu_{\text{core}})$, which render them exclusive w.r.t. emissions above q_{min}. To compute the Sudakov form factors Δ_t, we include quark masses in the splitting functions, according to the method described in [59, 60], using the corresponding extension of Eq. (3.2). This means in particular that we use the massive splitting functions from [61], the propagator corrections listed in [59, 60], and we replace the two-loop cusp term $K\, 2C_F/(1-z)$ by $K\, C_F(2/(1-z) - m^2/p_i p_j)$ in the case of massive quark splittings $\widetilde{ij} \to i, j$.

Scale uncertainties in the MiNLO framework are assessed through standard factor-two variations of μ_{R} and μ_{F}. The renormalisation scale is kept fixed in the Sudakov form factors but is varied as usual in the rest of the (N)LO cross section, including the counterterms that subtract the $\mathcal{O}(\alpha_{\text{s}})$ parts of the Sudakov form factors at NLO. Variations $\mu_{\text{F}} \to \xi_{\text{F}}\, \mu_{\text{F}}$ of the factorisation scale are more subtle. They have to be applied at the level of PDFs and related NLO counterterms, as well as in the Sudakov form factors that depend on $q_{\text{min}} = \mu_{\text{F}}$. More precisely, $q_{\text{min}} \to \xi_{\text{F}}\, q_{\text{min}}$ variations are applied only to Sudakov form factors associated with external and internal initial-state lines, and Sudakov form factors $\Delta_a(\xi_{\text{F}}\, q_{\text{min}}, q_k)$ are set to one when $\xi_{\text{F}}\, q_{\text{min}}$ exceeds q_k.



(a) (b)

Figure 3.1: Inclusive $t\bar{t}$+multijet cross sections with a minimum number $N = 0, 1, 2, 3$ of jets at $p_{\text{T,jet}} \geq 25\,\text{GeV}$ (a) and ratios of $t\bar{t} + N\,\text{jet}$ over $t\bar{t} + (N-1)\,\text{jet}$ inclusive cross sections for $N = 1, 2, 3$ and $p_{\text{T,jet}} \geq 25\,\text{GeV}$ (b). See the main text for details.

3.4 Predictions for the 13 TeV LHC

In the following we present selected predictions for $pp \to t\bar{t} + 0, 1, 2, 3\,\text{jets}$ at 13 TeV. We construct jets by clustering light partons with the anti-k_t algorithm [65] at $R = 0.4$, and by default we select jets with pseudorapidity $|\eta_{\text{jet}}| < 2.5$ and a jet-p_{T} threshold of 25 GeV. Unless stated otherwise, depending on the minimum number N of jets that is required by the observable at hand, inclusive (N)LO or MI(N)LO calculations with N jets are used.

The jet multiplicity distribution is presented in Fig. 3.1(b). The top panel displays four predictions, stemming from fixed-order LO and NLO calculations, and from MINLO computations at LO and NLO (labeled 'MILO' and 'MINLO'). The second panel shows the ratio between LO and NLO predic-

tions at fixed order, while the third panel shows the ratio between MILO and MINLO predictions. The last panel shows the ratio between MINLO and NLO. The bands illustrate scale uncertainties estimated through independent factor-two rescaling of μ_R and μ_F excluding antipodal variations. Fixed-order predictions feature rather large NLO corrections of about +50% for all jet multiplicities, while MiNLO results feature steadily decreasing corrections for increasing N_{jets}. In both cases, LO scale uncertainties tend to grow by more than 10% at each extra jet emission, while (MI)NLO scale uncertainties are significantly reduced and the total width of the (MI)NLO variation bands is about 20–25% for all considered N_{jets} values. Comparing fixed-order NLO and MINLO predictions we observe a remarkable agreement at the level of 4–8%. This supports NLO and MINLO scale-uncertainty estimates based on factor-two variations and encourages the usage of either of the two calculations (NLO and MINLO) in practical applications.

As demonstrated in Table 3.2, the good agreement between fixed-order NLO and MINLO results and the consistency of the observed NLO–MINLO differences with factor-two scale variations persist also for a range of other commonly used $p_{T,\text{jet}}$-thresholds [144]. More precisely, for inclusive $t\bar{t}+N$ jet cross sections with jet-p_T thresholds of 25, 40, 60 and 80 GeV, MINLO predictions lie systematically above NLO ones. The difference amounts to 6% in the inclusive case and grows up to 19% at large jet multiplicity and for large p_T-thresholds, in which case MiNLO cross sections feature significantly better perturbative convergence and smaller scale uncertainties as compared to fixed-order ones. This can be attributed to the improved treatment of multijet events with multiple scales through logarithmic resummation in MiNLO. In Table 3.2 also exclusive cross sections with exactly N jets are presented. In that case, the difference between MINLO and NLO predictions varies between -7% and +11%. Apart from the zero-jet case, where the MiNLO approach is not well motivated, the MINLO/NLO ratio is almost independent of the number of jets and grows from 0.95 to 1.10 when the p_T-threshold increases from 25 to 80 GeV. Similarly as in the inclusive case, at p_T-thresholds above 40 GeV MINLO predictions for exclusive N-jet cross sections with $N \geq 2$ feature much better convergence and smaller scale uncertainties w.r.t. fixed order. However, for lower p_T-thresholds the opposite is observed, and in the three-jet case the MINLO scale uncertainty becomes twice as large at the NLO one. This can be attributed to the fact that Sudakov logarithms related to the vetoing of NLO radiation are not resummed in the MiNLO approach. In spite of this caveat, the general agreement of fixed-order NLO and MiNLO results remains remarkably good and typically consistent with respective scale uncertainties. Actually, at large jet multiplicity and p_T, nominal MiNLO predictions can lie slightly outside NLO scale variations and/or

vice versa, but this tension never excees the 5% level, while MiNLO and NLO scale variations are found to overlap for all considered observables in Table 3.2.

Figure 3.1(a) shows ratios of inclusive $t\bar{t} + N$ jet cross sections for successive jet multiplicities. Due to the cancellation of various sources of experimental and theoretical uncertainties, such ratios are ideally suited for precision tests of QCD. Corresponding ratios have been widely studied in vector-boson plus multi-jet production [145, 146], where a striking scaling behavior was observed at high jet multiplicity. In the case of $t\bar{t}$+multijet ratios involving up to three jets we find a moderate dependence on the number of jets but no clear scaling. This behavior is rather similar to scaling violations in $V+$multijet production at lower multiplicity and, analogously as for $V+$multijets, can be attributed to the suppression of important partonic channels in the zero-jet process at LO. In fact, quark–gluon channels are not active in $t\bar{t}$ production at LO. In addition, at LHC energies the gluonic initial state is strongly favored due to the parton luminosity and the t-channel enhancement of the $gg \to t\bar{t}$ cross section, such that the situation becomes similar to vector boson production, except for the difference of quark versus gluon initial states at LO. When adding additional jets, firstly quark–gluon initial states and secondly quark–quark initial states (including t-channel top-quark diagrams) are added, which contribute sizably to the cross section at larger invariant mass and/or transverse momentum. In order to test scaling hypotheses, it would therefore ultimately be necessary to compute the $t\bar{t}+4$ jet over $t\bar{t}+3$ jet ratio, and eventually the $t\bar{t}+5$ jet over $t\bar{t}+4$ jet ratio. This is out of reach of present technology, therefore we do not investigate the scaling behavior in more detail. Nevertheless, given the excellent agreement between MINLO and NLO predictions up to three jets, the ratios in Fig. 3.1(a) can be regarded as optimal benchmarks for precision tests.

Figure 3.2 shows the transverse momentum spectrum of the top quark for varying jet multiplicities. From low to very high p_{T} NLO scale uncertainties remain at a similarly small level as for integrated cross sections. For $N_{\mathrm{jets}} \geq 1$, we observe significant shape corrections, which tend to decrease at high jet multiplicity in MiNLO, while in fixed order they remain important. We also observe a shape difference between fixed-order and MiNLO predictions, which tends to increase with increasing jet multiplicity but is clearly reduced at NLO. The overall agreement between fixed-order NLO and MINLO results is quite good, both in shape and normalisation, with differences that lie within the individual scale uncertainties. Figure 3.3 shows the top-quark pair transverse momentum spectrum in 1-, 2- and 3-jet samples. We observe a large increase in the cross section between LO and NLO in the one-jet case,

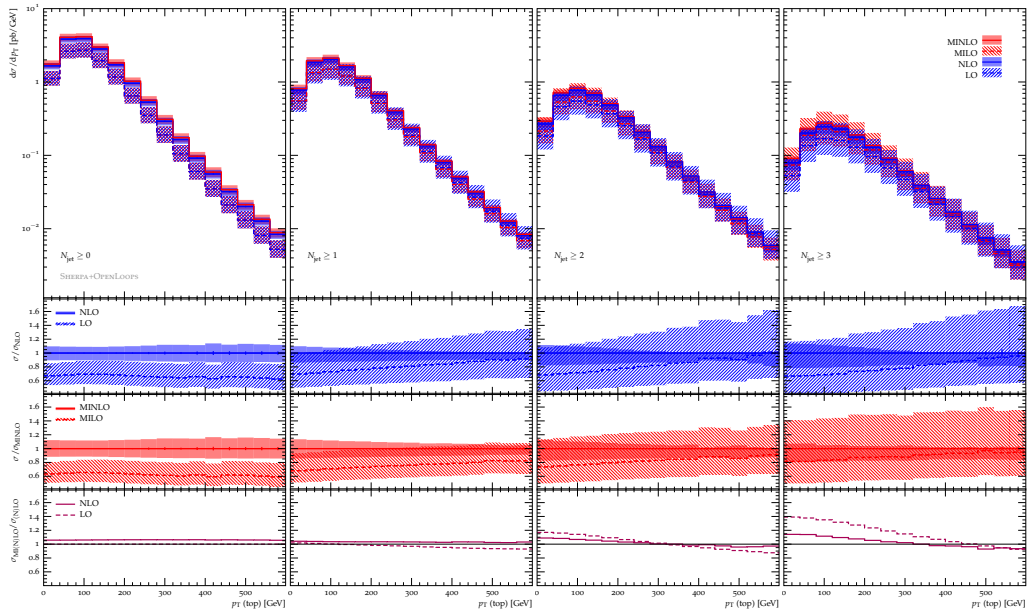


Figure 3.2: Distribution in the top-quark $p_{\rm T}$ for $pp \to t\bar{t} + 0, 1, 2, 3$ jets with $p_{\rm T,jet} \geq 25\,\text{GeV}$.

where the effect of additional radiation not modeled by the LO calculation is largest. At higher jet multiplicities correction effects tend to decrease. Fixed-order NLO uncertainties are similarly small as in Fig. 3.2, while MINLO scale uncertainties tend to be more pronounced in the tails. However, we find a very good overall agreement between fixed-order NLO and MINLO predictions, especially for $N_{\rm jets} \geq 2$ and 3.

The jet transverse momentum spectrum of the first, second and third jet, as predicted by $t\bar{t} + N$ jet calculations of corresponding jet multiplicity, is displayed in Fig. 3.4. In general we observe approximately constant NLO K-factors over the entire range of transverse momenta analyzed here, but in terms of perturbative convergence and scale uncertainties at NLO we find that the MiNLO approach performs better than fixed order. Comparing fixed-order and MiNLO results, at LO we find significant deviations that grow with $N_{\rm jets}$ and can reach 60% in the tails. Such differences are largely reduced by the transition to NLO. The fairly decent agreement between fixed-order NLO and MINLO results exemplifies nicely how the convergence of the perturbative series leads to a reduced dependence not only on constant scale variations, but also on the functional form of the scale.

Figure 3.5 shows inclusive $t\bar{t} + 1, 2, 3$ jet predictions for the total light-jet transverse energy, which is defined as $H_{\rm T}^{\rm jets} = \sum_j |p_{{\rm T},j}|$, with the sum

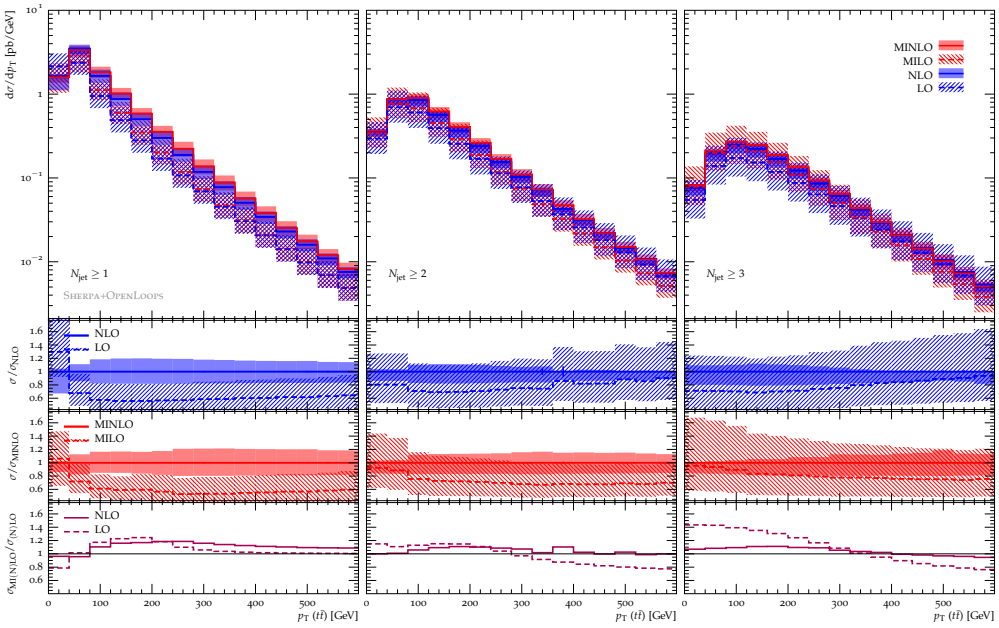


Figure 3.3: Distribution in the p_{T} of the $t\bar{t}$ system for $pp \to t\bar{t} + 1, 2, 3$ jets with $p_{\text{T,jet}} \geq 25\,\text{GeV}$.

running over all reconstructed jets within acceptance. This observable is typically badly described by LO calculations, as a sizable fraction of events, especially at large $H_{\text{T}}^{\text{jets}}$, contains additional jets originating in initial-state radiation [147]. Correspondingly we observe a very large increase in the cross section between LO and NLO in the one-jet samples, where the effect of additional radiation not modeled by the calculation is largest. At higher jet multiplicities, the increase is smaller, but well visible. In MiNLO it tends to be more pronounced than at fixed order, and for $N_{\text{jets}} \geq 3$ also MINLO uncertainties are larger than NLO ones. Nevertheless, we find good overall agreement between fixed-order NLO and MINLO predictions, independent of the jet multiplicity. However, given the strong sensitivity of $H_{\text{T}}^{\text{jets}}$ to multi-jet emissions, NLO or MINLO calculations with fixed jet multiplicity might significantly underestimate the effect of additional QCD radiation, and an approach like multijet merging at NLO [32] would be more appropriate for this particular observable.

Studying differential distributions in several angular variables we did not find any sizable shape effect. We thus refrain from showing corresponding plots.

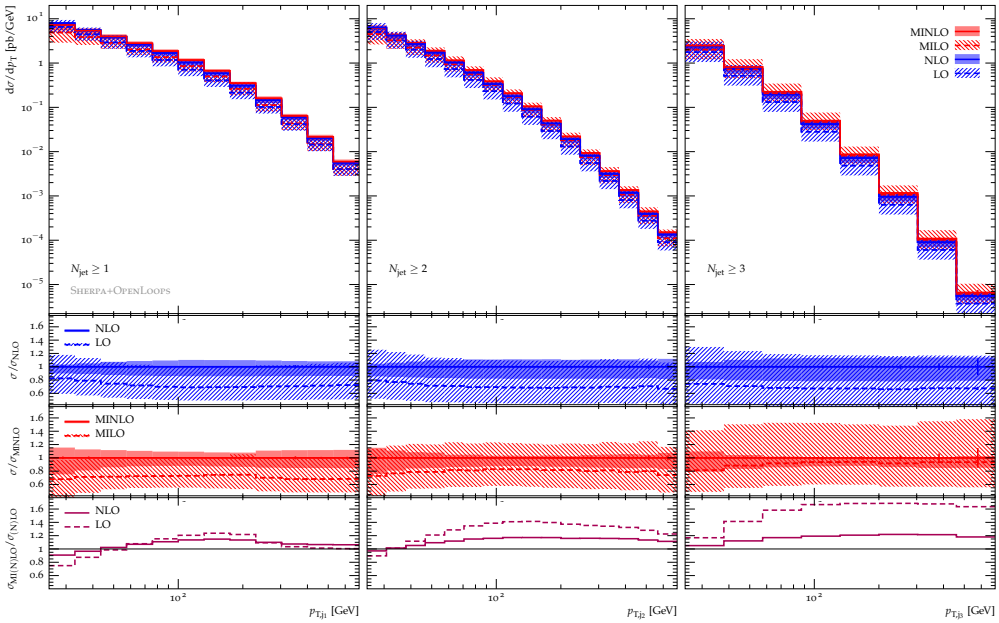


Figure 3.4: Distribution in the p_T of the n-th jet for $pp \to t\bar{t} + n\,\text{jets}$ with $p_{T,\text{jet}} \geq 25\,\text{GeV}$ and $n = 1, 2, 3$.

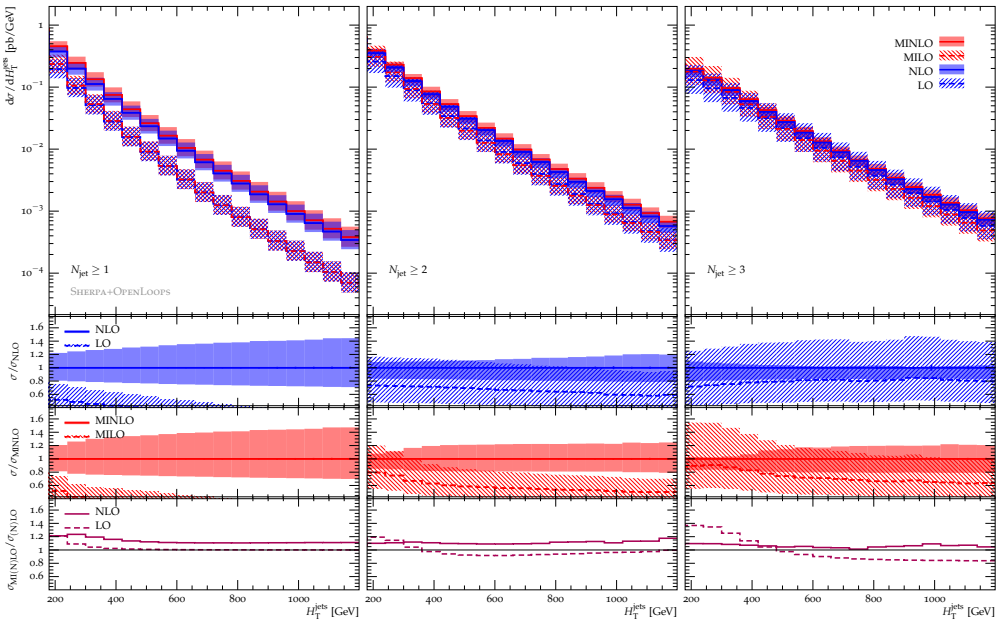


Figure 3.5: Distribution in the total transverse energy of light jets for $pp \to t\bar{t} + 1, 2, 3\,\text{jets}$ with $p_{T,\text{jet}} \geq 25\,\text{GeV}$.

3.5 Conclusions

We have computed predictions for top-quark pair production with up to three additional jets at the next-to-leading order in perturbative QCD using the automated programs OpenLoops and Sherpa. This is the first calculation of this complexity involving massive QCD partons in the final state. Given the multi-scale nature of $t\bar{t}$+multijet production, finding a scale that guarantees optimal perturbative convergence is not trivial. Moreover, standard factor-two scale variations might not provide a correct estimate of theoretical uncertainties related to missing higher-order effects. These issues have been addressed by comparing predictions obtained at fixed order using the scale $H_{\mathrm{T}}/2$ and, alternatively, with the MiNLO method. The hard scale $H_{\mathrm{T}}/2$ is known to yield good perturbative convergence for a large class of processes, while the MiNLO approach is more favorable from the theoretical point of view, as it implements NLL resummation for soft and collinear logarithms that emerge in the presence of large ratios of scales. For a rather wide range of observables at the 13 TeV LHC, we find very good agreement between the predictions generated at fixed order and with the MiNLO method. More precisely, MiNLO cross sections tend to feature a better perturbative convergence and to lie systematically above NLO ones. But the differences turn out to be well consistent with factor-two scale variations of the respective predictions, which are typically at the 10% level. These findings significantly extend and consolidate the conventional picture of theoretical uncertainties that results from standard scale choices and scale variations.

$p_{\mathrm{T,jet}} \geq 25\,\mathrm{GeV}$	LO	NLO	MILO	MINLO	$\frac{\text{MINLO}}{\text{NLO}}$	$\frac{\text{MINLO}}{\text{MILO}}$	$\frac{\text{NLO}}{\text{LO}}$
$N_{\mathrm{jets}} \geq 0$	$440.46(22)^{+28\%}_{-21\%}$	$644.34(31)^{+9\%}_{-11\%}$	$440.46(22)^{+28\%}_{-21\%}$	$683.21(31)^{+12\%}_{-12\%}$	1.06	1.55	1.46
$N_{\mathrm{jets}} = 0$		$375.36(41)^{+3\%}_{-15\%}$		$394.66(41)^{+2\%}_{-11\%}$	1.05	0.90	0.85
$N_{\mathrm{jets}} \geq 1$	$268.93(9)^{+43\%}_{-28\%}$	$361.01(26)^{+7\%}_{-13\%}$	$267.86(9)^{+36\%}_{-24\%}$	$373.94(23)^{+11\%}_{-13\%}$	1.04	1.40	1.34
$N_{\mathrm{jets}} = 1$		$249.52(27)^{+2\%}_{-17\%}$		$237.88(25)^{+5\%}_{-14\%}$	0.95	0.89	0.93
$N_{\mathrm{jets}} \geq 2$	$111.32(3)^{+59\%}_{-35\%}$	$149.43(16)^{+8\%}_{-16\%}$	$122.86(4)^{+53\%}_{-32\%}$	$157.49(16)^{+9\%}_{-16\%}$	1.05	1.28	1.34
$N_{\mathrm{jets}} = 2$		$111.11(16)^{+1\%}_{-18\%}$		$103.98(17)^{+5\%}_{-21\%}$	0.94	0.85	1.00
$N_{\mathrm{jets}} \geq 3$	$38.36(2)^{+75\%}_{-40\%}$	$53.01(10)^{+9\%}_{-20\%}$	$48.89(2)^{+71\%}_{-38\%}$	$57.43(11)^{+4\%}_{-18\%}$	1.08	1.17	1.38
$N_{\mathrm{jets}} = 3$		$41.12(10)^{+0\%}_{-19\%}$		$38.14(12)^{+5\%}_{-38\%}$	0.93	0.78	1.07
$p_{\mathrm{T,jet}} \geq 40$ GeV	LO	NLO	MILO	MINLO	$\frac{\text{MINLO}}{\text{NLO}}$	$\frac{\text{MINLO}}{\text{MILO}}$	$\frac{\text{NLO}}{\text{LO}}$
$N_{\mathrm{jets}} \geq 0$	$440.46(22)^{+28\%}_{-21\%}$	$644.34(31)^{+9\%}_{-11\%}$	$440.46(22)^{+28\%}_{-21\%}$	$683.21(31)^{+12\%}_{-12\%}$	1.06	1.55	1.46
$N_{\mathrm{jets}} = 0$		$461.03(36)^{+1\%}_{-4\%}$		$483.40(36)^{+1\%}_{-5\%}$	1.05	1.10	1.05
$N_{\mathrm{jets}} \geq 1$	$183.17(7)^{+44\%}_{-28\%}$	$255.88(20)^{+9\%}_{-13\%}$	$200.36(7)^{+35\%}_{-24\%}$	$276.63(20)^{+10\%}_{-12\%}$	1.08	1.38	1.40
$N_{\mathrm{jets}} = 1$		$201.57(20)^{+0\%}_{-8\%}$		$206.41(21)^{+3\%}_{-7\%}$	1.02	1.03	1.10
$N_{\mathrm{jets}} \geq 2$	$54.23(2)^{+59\%}_{-35\%}$	$76.13(8)^{+10\%}_{-17\%}$	$68.34(2)^{+51\%}_{-31\%}$	$84.71(10)^{+6\%}_{-14\%}$	1.11	1.24	1.40
$N_{\mathrm{jets}} = 2$		$62.30(8)^{+0\%}_{-12\%}$		$63.60(10)^{+2\%}_{-14\%}$	1.02	0.93	1.15
$N_{\mathrm{jets}} \geq 3$	$13.84(1)^{+75\%}_{-40\%}$	$19.87(4)^{+11\%}_{-21\%}$	$20.55(1)^{+68\%}_{-37\%}$	$22.70(6)^{+2\%}_{-15\%}$	1.14	1.10	1.44
$N_{\mathrm{jets}} = 3$		$16.61(4)^{+1\%}_{-16\%}$		$16.80(6)^{+0\%}_{-33\%}$	1.01	0.82	1.20
$p_{\mathrm{T,jet}} \geq 60$ GeV	LO	NLO	MILO	MINLO	$\frac{\text{MINLO}}{\text{NLO}}$	$\frac{\text{MINLO}}{\text{MILO}}$	$\frac{\text{NLO}}{\text{LO}}$
$N_{\mathrm{jets}} \geq 0$	$440.46(22)^{+28\%}_{-21\%}$	$644.34(31)^{+9\%}_{-11\%}$	$440.46(22)^{+28\%}_{-21\%}$	$683.21(31)^{+12\%}_{-12\%}$	1.06	1.55	1.46
$N_{\mathrm{jets}} = 0$		$521.32(33)^{+3\%}_{-6\%}$		$547.02(33)^{+5\%}_{-8\%}$	1.05	1.24	1.18
$N_{\mathrm{jets}} \geq 1$	$123.16(5)^{+44\%}_{-29\%}$	$175.69(15)^{+9\%}_{-14\%}$	$142.02(6)^{+35\%}_{-24\%}$	$195.17(17)^{+9\%}_{-12\%}$	1.11	1.37	1.43
$N_{\mathrm{jets}} = 1$		$149.59(16)^{+3\%}_{-10\%}$		$160.41(17)^{+2\%}_{-8\%}$	1.07	1.13	1.21
$N_{\mathrm{jets}} \geq 2$	$26.06(1)^{+59\%}_{-35\%}$	$37.47(5)^{+11\%}_{-18\%}$	$35.24(2)^{+49\%}_{-30\%}$	$43.02(8)^{+4\%}_{-13\%}$	1.15	1.22	1.44
$N_{\mathrm{jets}} = 2$		$32.50(5)^{+4\%}_{-14\%}$		$35.10(8)^{+2\%}_{-9\%}$	1.08	1.00	1.25
$N_{\mathrm{jets}} \geq 3$	$4.95(0)^{+74\%}_{-40\%}$	$7.31(2)^{+13\%}_{-22\%}$	$7.97(1)^{+65\%}_{-36\%}$	$8.61(3)^{+1\%}_{-13\%}$	1.18	1.08	1.48
$N_{\mathrm{jets}} = 3$		$6.41(2)^{+5\%}_{-18\%}$		$6.89(3)^{+1\%}_{-27\%}$	1.07	0.86	1.29
$p_{\mathrm{T,jet}} \geq 80$ GeV	LO	NLO	MILO	MINLO	$\frac{\text{MINLO}}{\text{NLO}}$	$\frac{\text{MINLO}}{\text{MILO}}$	$\frac{\text{NLO}}{\text{LO}}$
$N_{\mathrm{jets}} \geq 0$	$440.46(22)^{+28\%}_{-21\%}$	$644.34(31)^{+9\%}_{-11\%}$	$440.46(22)^{+28\%}_{-21\%}$	$683.21(31)^{+12\%}_{-12\%}$	1.06	1.55	1.46
$N_{\mathrm{jets}} = 0$		$555.85(32)^{+5\%}_{-8\%}$		$584.21(32)^{+7\%}_{-9\%}$	1.05	1.33	1.26
$N_{\mathrm{jets}} \geq 1$	$88.46(4)^{+44\%}_{-29\%}$	$127.22(12)^{+10\%}_{-14\%}$	$104.19(5)^{+34\%}_{-25\%}$	$142.99(14)^{+9\%}_{-12\%}$	1.12	1.37	1.44
$N_{\mathrm{jets}} = 1$		$112.89(12)^{+5\%}_{-12\%}$		$123.77(14)^{+3\%}_{-9\%}$	1.10	1.19	1.28
$N_{\mathrm{jets}} \geq 2$	$14.33(1)^{+59\%}_{-35\%}$	$20.81(3)^{+11\%}_{-18\%}$	$19.90(1)^{+48\%}_{-30\%}$	$24.22(4)^{+3\%}_{-12\%}$	1.16	1.22	1.45
$N_{\mathrm{jets}} = 2$		$18.64(3)^{+6\%}_{-15\%}$		$20.71(4)^{+1\%}_{-9\%}$	1.11	1.04	1.30
$N_{\mathrm{jets}} \geq 3$	$2.17(0)^{+74\%}_{-40\%}$	$3.22(1)^{+13\%}_{-22\%}$	$3.59(0)^{+63\%}_{-36\%}$	$3.85(2)^{+1\%}_{-13\%}$	1.19	1.07	1.48
$N_{\mathrm{jets}} = 3$		$2.91(1)^{+8\%}_{-19\%}$		$3.23(2)^{+1\%}_{-23\%}$	1.11	0.90	1.34

Table 3.2: Inclusive ($N_{\mathrm{jets}} \geq n$) and exclusive ($N_{\mathrm{jets}} = n$) cross sections with $n = 0, 1, 2, 3$ jets and different transverse momentum thresholds, $p_{\mathrm{T,jet}} \geq 25, 40, 60, 80$ GeV. Uncertainties represent the envelope of the independent μ_{R} and μ_{F} variations around the central value (antipodal variations excluded).

Summary and Conclusions

As the heaviest fundamental particle in the Standard Model, with a mass of order of the electroweak scale, the top quark is widely believed to play a special role in the mechanism of electroweak symmetry breaking and to be connected to physics Beyond the Standard Model. In this context, the production of a Higgs boson in association with a top-antitop pair can provide a unique test of the mechanism that generates the top-quark mass, as it is the only Higgs-production mode which allows for a direct measurement of the top Yukawa coupling. Among the various Higgs-decay modes that can be exploited for $t\bar{t}H$ searches, the $H \to b\bar{b}$ channel offers the largest branching ratio. At the same time, a measurement of the $t\bar{t}H(b\bar{b})$ cross section would give simultaneous access to the top and bottom Yukawa couplings. However, this channel is plagued by overwhelming QCD backgrounds that have hampered its experimental observation so far. In particular, theoretical uncertainties in the simulation of the QCD backgrounds represent the main source of uncertainty in $t\bar{t}H(b\bar{b})$ searches.

In this work we have presented precise theoretical predictions for the dominant QCD backgrounds to $t\bar{t}H(b\bar{b})$ production, namely top-pair production in association with multiple jet emissions ($pp \to t\bar{t}$+multijets), in particular when a $t\bar{t}$ system is produced in association with a $b\bar{b}$ pair ($pp \to t\bar{t}b\bar{b}$). The calculations for the relevant multi-particle processes, at NLO QCD, have been performed using the Sherpa Monte Carlo framework in conjunction with the one-loop matrix-element generator OpenLoops. Methodological aspects concerning fixed-order NLO calculations, their matching to parton showers, the multi-jet merging technique, and the MINLO approach, have been reviewed in detail in Chapter one.

In Chapter two we have presented NLO QCD predictions for $t\bar{t}b\bar{b}$ production at the LHC. This channel corresponds to the irreducible background to the $t\bar{t}H(b\bar{b})$ signal. Our NLO predictions include b-quark mass effects and are matched to the Sherpa parton shower using the S-Mc@NLO method. This allowed us to extend NLO predictions to arbitrary $t\bar{t}b\bar{b}$ kinematics, including the case where one or both b–jets arise from collinear $g \to b\bar{b}$ splittings. We

found that this splitting mechanism plays an important role for the $t\bar{t}H(b\bar{b})$ analysis, with effects of order +20–30% in the Higgs signal region.

The production of $t\bar{t}$ pairs in association with multiple light jets represents another large background to $t\bar{t}H$ production in the $H \to b\bar{b}$ channel. Experimental analyses of $t\bar{t}H(b\bar{b})$ exploit several event categories with varying number of light jets and b-jets. This fact, together with the possibility to misidentify experimentally light jets as b-jets, leads to the need of very accurate and realistic simulations of top-pair production in combination with both light and b-flavoured jets. In the third chapter, we have presented theoretical predictions for the production of top-quark pairs with up to three jets at NLO QCD. This is the first calculation that deals with a $2 \to 5$ process with seven coloured external particles including also heavy quarks. To address the issue of scale choices and related uncertainties in the presence of multiple scales, we compared results obtained with a fixed order calculation and the MiNLO procedure. We found that the differences are typically below the respective factor-two scale variations, suggesting that for all considered jet multiplicities missing higher-order effects should not exceed the ten percent level.

The results presented in this thesis represent the current state of the art in the precise description of $t\bar{t}H(b\bar{b})$ backgrounds. At the same time, they should be regarded as a first important step towards a unified description of $t\bar{t}$+multijet production including light and heavy-flavour jets.

Acknowledgements

And here we are, finally, at the end of this thesis, which closes these four years of work. To be honest, I'm a bit sad to leave this institute. These four years here have been wonderful, thanks also to all the people I came across in this period. Maybe most of the people will not read the manuscript this far, or maybe somebody will open it and read only this part. Or maybe there will be people who will not even open this thesis at all. Well, anyway, this is the time for the acknowledgements I need to do.

The first acknowledgement goes to Stefano, my supervisor. Thank you for giving me this opportunity. You took a young and inexpert student and you made him a not-so-young and expert (well... let's be humble and say "less inexpert") PhD-student. I want to thank also all the colleagues that helped me in these years in several and disparate projects. Thanks to Philipp, who was maybe the first real programmer I've worked with. Your "why-writing-50-lines-of-code-if-you-can-shrink-in-a-cryptic-long-one" philosophy has been enlightening for me. Maybe it was a bit less when I had to go through your code, but... I thank Stefan and Jonas for the help during our projects. I thank Marek for the patience he had all the times I knocked on his door for help. My gratitude goes also to all my former and current officemates. Thank you Fabio, James, Graeme, Matthieu, Max and Federico, for having made our office 36K42 a very funny and enjoyable place. I'm also grateful to our secretaries, Esther and Regina, for the vital help they gave me with the bureaucracy and, not less important, with all the stuff in German I got. In general, I want to thank all the "Institut für Theoretische Physik" for such a nice and stimulating environment during these years. I think it will be very hard to find a better place to work.

I'm also grateful to all the friends I have outside the institute. I want to thank Fabio, Lorenzo and their ex-flatmate Davide, for having been the first friends when I arrived, when I didn't know anyone here. Thanks to Davide, Michela, Chiara, Felicity, Roberto and all the people of the *tabletop society*, for the funny evenings and delicious dinners we had (and still having) together.

Finally thanks to my family, grazie ai miei genitori per avermi sempre supportato (e sopportato) incondizionatamente in tutte le scelte che ho fatto finora, anche se mi hanno portato lontano. Grazie a mia sorella e mio fratello per l'affetto durante questi anni, perché ogni volta che torno a casa dopo una lunga assenza, sembra che non me ne sia mai veramente andato. Un grazie particolare va anche al nostro cane Rebus, che mi accoglie a festa ad ogni mio ritorno (anche se sospetto che sia perché pensa che abbia del cibo nelle valigie...).

Last but not least a special thank goes to my girlfriend Valentina. Grazie per essere stata al mio fianco durante questi anni, e per continuare ad esserci. Grazie per avermi dato un motivo per smontare dal lavoro e ritornare a casa ogni giorno. Grazie di cuore perché senza di te tutta questa esperienza sarebbe stata piú spenta.

Appendix A

Branching probabilities and Sudakov form factors in MiNLO

In this appendix we document the branching probabilities that have been used in the definitions of the Sudakov form factors of Eqs. (1.45), (1.47) and (1.49) in MiNLO and in Chapter 3. Let us study the splitting

$$Q \to Q(p_1) + g(p_2),$$

with Q being a quark of mass m. The extension to the other splittings can be done in a similar way. We can express the momenta using the Sudakov parametrisation [59]

$$p_1^\mu = z p^\mu - q_\perp^\mu + \frac{\mathbf{q}^2 + p_1^2}{z} \frac{n^\mu}{2n \cdot p}, \tag{A.1}$$

$$p_2^\mu = (1-z) p^\mu + q_\perp^\mu + \frac{\mathbf{q}^2 + p_2^2}{1-z} \frac{n^\mu}{2n \cdot p}, \tag{A.2}$$

where p^μ and n^μ are light-like vectors: $p^2 = n^2 = 0$, and $q_\perp^\mu$ is the space-like transverse momentum, with $p \cdot q_\perp = n \cdot q_\perp = 0$, with $\mathbf{q}^2 = -q_\perp^2$.

In the collinear limit, the $n+1-$parton matrix element M_{n+1} fulfils a factorisation formula, and it can be related to the M_n matrix element through the relation

$$|M_{n+1}(p_1, p_2, \dots)|^2 \simeq |M_n(p_1 + p_2, \dots)|^2 \frac{4\pi\alpha_S}{p_1 \cdot p_2} P_{QQ}(z, q) \tag{A.3}$$

$$\simeq |M_n(p_1 + p_2, \dots)|^2 8\pi\alpha_S \frac{z(1-z)}{\mathbf{q}^2 + (1-z)^2 m^2} P_{QQ}(z, q) \tag{A.4}$$

where $P_{QQ}(z,q)$ is the (unregularized) spin-averaged splitting function

$$\begin{aligned} P_{QQ}(z,q) &= \mathrm{C_F}\left[\frac{1+z^2}{1-z} - \frac{m^2}{p_1\cdot p_2}\right] \\ &= \mathrm{C_F}\left[\frac{1+z^2}{1-z} - \frac{2z(1-z)m^2}{\mathbf{q}^2+(1-z)^2m^2}\right]. \end{aligned}$$

Similar collinear factorisation relations hold also for the other splittings. The starting point is the definition of the Sudakov form factor given in (1.14), that is the probability of a parton of specie a to not emit any parton between the scales t_0, t_1,

$$\Delta_a(t_0,t_1) = \exp\left\{-\int_{t_0}^{t_1} dt \frac{\alpha_S(t)}{2\pi t}\sum_{b=q,g}\int_{z_0}^{1-z_0} dz P_{ab}(z) + \delta_{ab}\frac{2C_a}{1-z}\Lambda_2(t)\right\}, \tag{A.5}$$

where $a = q, g$ is the parton specie, $P_{ab}(z)$ are the splitting functions related to splitting $a \to bc$, $z_0 = \sqrt{t/t_1}$ and $t = \mathbf{q}^2$ is the (squared) relative p_T of the splitting, as defined in Eqs. (A.1) and (A.2). Note that in our implementation we include also the two-loop correction [51] to the DGLAP splitting functions

$$\Lambda_2(t) = \frac{\alpha_S(t)}{2\pi}\left[\left(\frac{67}{18} - \frac{\pi^2}{6}\right)\mathrm{C_A} - \frac{10}{9}\mathrm{T_R}\, N_f(t)\right], \tag{A.6}$$

where $C_a = \mathrm{C_F}, \mathrm{C_A}$ are the group theoretical constants for $a = q, g$ respectively, and $N_f(t)$ is the number of the active massless flavours at the scale t.

A.1 Massless quarks

Here we will consider the massless splitting $q \to qg$. Branching probabilities used in this work have been obtained as follows. Starting from the LO DGLAP massless splitting function $P_{qq}(z)$, we use the trivial identity

$$P_{qq}(z) = zP_{qq}(z) + (1-z)P_{qg}(1-z),$$

where P_{qg} is the splitting function associated with the $q \to gq$ splitting, and $P_{qq}(z) = P_{qg}(1-z)$. The integral over the momentum fraction z can then be

rewritten as

$$\Gamma_{qq}(\epsilon) = \int_{z-}^{z+} dz P_{qq}(z) = \int_{z-}^{z+} dz(zP_{qq}(z) + (1-z)P_{qg}(1-z)) \tag{A.7}$$

$$= \int_{z-}^{z+} dzz(P_{qq}(z) + P_{qg}(z)), \tag{A.8}$$

where $z- = \epsilon, z+ = 1-\epsilon$ are boundaries symmetric with respect to $1/2$ and $\epsilon = sqrtt/t_1$. In the last line we made use of the symmetry $z \leftrightarrow 1-z$ of the quark-gluon splitting function. Finally we can extend the integration lower bound $z- \to 0$, since the difference from the full result is a power correction of $O(\epsilon)$. This approach guarantees the branching probabilities to exist over all the phase space an to be positive across all the integration domain, allowing the interpretation of Sudakov form factors in terms of no-branching probabilities.

The procedure outlined can be summarized in the following formula for the massless Sudakovs FFs:

$$\Delta_a(t_0, t_1) = \exp\left\{ - \int_{t_0}^{t_1} \frac{dt}{t} \frac{\alpha_S(t)}{2\pi} \sum_{b=q,g} \int_0^{1-\sqrt{t/t_1}} \mathrm{d}z \left(z\, P_{ab}(z) + \delta_{ab} \frac{2C_a}{1-z} \Lambda_2(t) \right) \right\} . \tag{A.9}$$

The explicit expressions for the Sudakov FFs of a given specie $a = q, g$ are

$$\Delta_q(t_0, t_1) = \exp\left[- \int_{t_0}^{t_1} \frac{dt}{t} \frac{\alpha_S(t)}{2\pi} \Gamma_{qq}\left(\sqrt{\frac{t}{t_1}} \right) \right], \tag{A.10}$$

$$\Delta_g(t_0, t_1) = \exp\left[- \int_{t_0}^{t_1} \frac{dt}{t} \frac{\alpha_S(t)}{2\pi} \left(\Gamma_{gg}\left(\sqrt{\frac{t}{t_1}} \right) + N_f(t) \Gamma_{gq}\left(\sqrt{\frac{t}{t_1}} \right) \right) \right], \tag{A.11}$$

where the branching probabilities Γ_{ab} in the massless case read

$$\Gamma_{gg}(\epsilon) = \mathrm{C_A} \left[2(\Lambda_2 + 1) \log\left(\frac{1}{\epsilon} \right) - \frac{1}{6}(1-\epsilon)^2 (11 - (2-3\epsilon)\epsilon) \right], \tag{A.12}$$

$$\Gamma_{qq}(\epsilon) = \mathrm{C_F} \left[2(\Lambda_2 + 1) \log\left(\frac{1}{\epsilon} \right) - \frac{3}{2}(1-\epsilon)^2 \right], \tag{A.13}$$

$$\Gamma_{gq}(\epsilon) = \frac{1}{3} \mathrm{T_R} (1-\epsilon)^2 \left(3\epsilon^2 + 2(1-\epsilon) \right), \tag{A.14}$$

with $\mathrm{T_R} = 1/2, \mathrm{C_F} = N_c$ and $\mathrm{C_A} = (N_c^2 - 1)/2N_c$ being the group theoretical constants. The Sudakov FFs (A.10) and (A.11) are then integrated numerically in our implementation, using the α_S routines provided by the PDFs.

A.2 Massive quarks

In case of massive quarks Q, the results are obtained following (A.9) using the massive LO splitting functions (see e.g. [59]) and replacing the massless propagator term with the massive one before performing the integration over the momentum fraction z, that is, with the replacements

$$\frac{1}{q^2} \to \frac{1}{q^2+m^2} \qquad \text{for } P_{gQ} \tag{A.15}$$

$$\left.\begin{aligned} &\frac{1}{q^2} \to \frac{1}{q^2+(1-z)^2m^2} \\ &\Lambda_2\, C_F \frac{2}{(1-z)} \to \Lambda_2\, C_F \left(\frac{2}{1-z} - \frac{2z(1-z)m^2}{q^2+(1-z)^2m^2} \right) \end{aligned}\right\} \qquad \text{for } P_{QQ} \tag{A.16}$$

The mass terms can be conveniently parametrised with the variable $\kappa \equiv q/m$. Repeating the derivation of (A.1) with these substitutions the Sudakov FFs in the massive case become

$$\Delta_Q(t_0,t_1) = \exp\left[-\int_{t_0}^{t_1} \frac{dt}{t} \frac{\alpha_S(t)}{2\pi} \Gamma_{QQ}\left(\sqrt{\frac{t}{t_1}}, \sqrt{\frac{t}{m^2}} \right) \right], \tag{A.17}$$

$$\begin{aligned} \Delta_g(t_0,t_1) = \exp\Bigg[-\int_{t_0}^{t_1} \frac{dt}{t} \frac{\alpha_S(t)}{2\pi} \Bigg\{ &\Gamma_{gg}\left(\sqrt{\frac{t}{t_1}} \right) + N_f(t)\, \Gamma_{gq}\left(\sqrt{\frac{t}{t_1}} \right) \\ &+ N_F(t)\, \Gamma_{gQ}\left(\sqrt{\frac{t}{t_1}}, \sqrt{\frac{t}{m^2}} \right) \Bigg\} \Bigg], \end{aligned} \tag{A.18}$$

where the massive branching probabilities read

$$\begin{aligned} \Gamma_{QQ}(\epsilon,\kappa) = \Gamma_{qq}(\epsilon) + \mathrm{C_F} \Bigg[&\epsilon(1-\epsilon)^2 \left(\frac{\epsilon}{\epsilon^2+\kappa^2} - \frac{1-\epsilon}{(1-\epsilon)^2+\kappa^2} \right) \\ &- \left(1 - \frac{\kappa^2}{2}\right) \left(\log\left(\frac{\kappa^2+1}{\epsilon^2+\kappa^2} \right) - 2\kappa \tan^{-1}\left(\frac{\epsilon}{\kappa} \right) \right) + \frac{1}{2}\left(1-\epsilon^2\right) \\ &- \kappa \left(\left(1-\kappa^2\right) \tan^{-1}\left(\frac{\epsilon\kappa}{-\epsilon+\kappa^2+1} \right) + \tan^{-1}\left(\frac{1}{\kappa} \right) \right) \\ &+ \Lambda_2 \left(\frac{\kappa^2(\epsilon-1)}{\epsilon^2+\kappa^2} + \kappa \tan^{-1}\left(\frac{(1-\epsilon)\kappa}{\epsilon+\kappa^2} \right) + \log\left(\frac{\epsilon^2+\kappa^2}{1+\kappa^2} \right) \right) \Bigg], \end{aligned} \tag{A.19}$$

$$\Gamma_{gQ}(\epsilon,\kappa) = \mathrm{T_R} \frac{\kappa^2}{\kappa^2+1} (1-\epsilon)^2 \left(1 - \frac{\kappa^2}{\kappa^2+1} \frac{(1-\epsilon)(3\epsilon+1)}{3} \right), \tag{A.20}$$

with $N_F(t)$ being the number of active massive flavours at the scale t. The last line of (A.19) accounts for the massive correction to the cusp term (A.6).

Again, the Sudakov FFs (A.17) and (A.18) are then integrated numerically in our implementation, using the α_S routines provided by the PDFs.

Appendix B

NLO+PS simulations of $t\bar{t}b\bar{b}$ production

In this appendix we report the systematic comparison between Monte Carlo generators for $t\bar{t}b\bar{b}$ production at the 13 TeV LHC, to appear in the fourth Yellow Report of the LHC Higgs Cross Section Working Group [122]. The aim of this comparison is to quantify the differences arising from various NLO+PS simulations based on different parton showers, matching schemes, and flavour number schemes, in the perspective of the assessment of the theoretical uncertainties related to this process. The study presented in the following is also the starting point of the analysis presented in Section 2.3.

B.1 Introduction

The production of $t\bar{t}$ pairs in association with two b-jets constitutes a large irreducible background to $t\bar{t}H$ production in the $H \to b\bar{b}$ channel, and the rather large uncertainty of Monte Carlo simulations of $t\bar{t}$ + b-jets production is one of the main limitations of current $t\bar{t}H(b\bar{b})$ searches at the LHC. A reliable theoretical description of $t\bar{t}$ production in association with two b-jets requires hard-scattering cross sections for the relevant partonic processes $q\bar{q}/gg \to t\bar{t}b\bar{b}$ at NLO QCD [33–35]. The inclusion of NLO QCD effects reduces scale uncertainties from the 70–80% level at LO to about 20–30%. To become applicable in the context of experimental analyses, NLO calculations need to be matched to parton showers. A NLO+PS simulation of $t\bar{t}b\bar{b}$ production based on the five flavour number scheme (5FNS), where b-quarks are treated as massless partons, was presented in [148, 149], while an alternative

NLO+PS simulation that includes b-quark mass effects in the four flavour number scheme (4FNS) was published in [37].

Finite b-quark masses permit to extend $t\bar{t}b\bar{b}$ matrix elements to the full phase space, including regions where b-quark pairs become collinear and matrix elements with $m_b = 0$ would be divergent. Thus, using $t\bar{t}b\bar{b}$ matrix elements with $m_b > 0$ in the 4FNS it is possible to simulate $t\bar{t} + b$-jets production in a fully inclusive way, including also signatures where a b-quark remains unresolved and a single b-jet is observed.[1] In contrast, the applicability of 5FNS calculations is limited to phase-space regions where the two b-quarks in $t\bar{t}b\bar{b}$ matrix elements have sufficient transverse momentum and angular separation in order to avoid the breakdown of the $m_b = 0$ approximation in the collinear regions. Such a requirement needs to be imposed at the level of generation cuts, i.e. before matching matrix elements to the parton shower, and one might expect that the resulting NLO+PS predictions for observables with two or more hard b-jets should be insensitive to generation cuts. However, this is not the case, since events with multiple hard b-jets can result from collinear $g \to b\bar{b}$ splittings in $t\bar{t}b\bar{b}$ matrix elements combined with the conversion of hard gluons into b-jets via $g \to b\bar{b}$ parton shower splittings. In fact, as pointed out in [37], this so-called double-splitting mechanism can lead to a sizable enhancement of the $t\bar{t}b\bar{b}$ background in the Higgs-signal region, $M_{bb} \sim M_H$.

Given the importance of (quasi) collinear $g \to b\bar{b}$ splittings, the choice of the flavour number scheme and the inclusion of b-quark mass effects play a critical role. For what concerns 5FNS calculations, while it is clear that setting $m_b = 0$ and omitting the singular phase-space regions leads to a logarithmic sensitivity to the unphysical generation cuts for observables with a single hard b-jet, double-splitting (or multiple-splitting) contributions imply such a sensitivity also for observables with two or more hard b-jets. Such a logarithmic dependence can naturally be avoided in the framework of NLO merging [20, 24, 150], where the singular phase space regions, defined in terms of an appropriate merging cut, are populated by the parton shower combined with matrix elements for $t\bar{t} + 0, 1$ jet production. However, applying NLO merging to $t\bar{t} + 0, 1, 2$ jet production [32] is technically much more challenging as compared to NLO+PS simulations of $t\bar{t}b\bar{b}$ production. Moreover, in the merging approach all b-quarks produced via double-splitting contributions would tend to arise from the parton shower, which implies a strong dependence on parton-shower modelling. In contrast, 4FNS simulations with $m_b > 0$ have the advantage that one of the $g \to b\bar{b}$ splittings is

[1] Here, and in the following, we consistently exclude top-decay products from the counting of b jets.

entirely described in terms of $t\bar{t}b\bar{b}$ matrix elements at NLO, while the additional hard b-jet arises from $t\bar{t}b\bar{b}g$ tree amplitudes matched to the parton shower via $g \to b\bar{b}$ splittings that can take place at any stage of the shower evolution. In conventional 4FNS calculations as the ones presented in this study, the number of active quark flavours is limited to four both in the evolution of the PDFs and α_s. Thus, renormalisation group logarithms associated with b- and t-quark loops are included only at fixed-order NLO.[2] As discussed in [37], such logarithms can be easily resummed by using modified 4FNS PDF sets that include all relevant quark flavours in the running of α_s. At $\sqrt{s} = 8\,\text{TeV}$ it was found that higher-order contributions of this type increase the NLO 4FNS $t\bar{t}b\bar{b}$ cross section by about 9% [37], while in the 5FNS they are naturally included.

Finally, we observe that simulations of $t\bar{t}b\bar{b}$ production in the 4FNS can be combined with fully inclusive $t\bar{t}$+jets samples based on the 5FNS in a rather straightforward way [77]. In fact, in order to avoid the double counting of b-quark production in the $t\bar{t}b\bar{b}$ and $t\bar{t}$+jets sample it is sufficient to veto events that involve b quarks in the $t\bar{t}$+jets sample. This prescription has to be applied after parton showering, and the b-quark veto should be restricted to showered $t\bar{t}$+jets matrix elements before top decays, i.e. it should not be applied to b quarks that arise from (showered) top decays or from the underlying event.

From the above discussion it is clear that the parton shower and the choice of the flavour number scheme play a critical role in the description of $t\bar{t}b\bar{b}$ production, and the thorough understanding of the related uncertainties is of prime importance for the success of $t\bar{t}H(b\bar{b})$ searches. As a first step in this direction, in the following we present a systematic comparison of various NLO+PS simulations of $t\bar{t}b\bar{b}$ production based on different parton showers, matching schemes, and flavour number schemes.

B.2 NLO+PS tools and simulations

Three different NLO+PS simulations of $t\bar{t}b\bar{b}$ production at $\sqrt{s} = 13\,\text{TeV}$ based on Sherpa+OpenLoops [28, 71, 74], MadGraph5_aMC@NLO+Pythia8 [27,151–153] and PowHel+Pythia8 [36, 152–154] have been compared. The various NLO matching methods, parton showers and flavour number schemes employed in the three simulations are

[2] In the 4FNS, b-quark contributions to the running of α_s are consistently restored at NLO accuracy by including b-quark loops in the matrix elements and renormalising them via zero-momentum subtraction. Also top-quark loop contributions to α_s have been renormalized via zero-momentum subtraction in the 4FNS.

summarised in Table B.1. The SHERPA+OPENLOOPS and MADGRAPH5_AMC@NLO simulations employ the 4FNS and the MC@NLO matching method[3] to combine matrix elements with the SHERPA and PYTHIA8 parton showers, respectively. Therefore possible differences between SHERPA+OPENLOOPS and MADGRAPH5_AMC@NLO predictions can be attributed to parton shower effects or to differences in the two implementations of the MC@NLO approach and related technical parameters. Instead, the POWHEL simulation differs from the other two simulations in at least two aspects: it is performed in the 5FNS and it employs the POWHEG matching method [16, 17]. Moreover, when comparing SHERPA+OPENLOOPS and POWHEL predictions one should keep in mind that also the respective parton showers (SHERPA and PYTHIA8) are different.

Since the POWHEL simulation is performed with massless b quarks,[4] in order to avoid collinear $g \to b\bar{b}$ singularities, the hard matrix elements need to be restricted to phase space regions where both b quarks remain resolved. This is achieved through the generation cuts

$$m_{bb} > 2\xi_m m_b, \qquad p_{T,b} > \xi_T m_b. \tag{B.1}$$

Technically, in order to guarantee infrared safety, in the case of real emission events the cuts in Eq. (B.1) are applied after a projection onto the Born phase space. The above cuts are chosen in a way that mimics, at LO in α_S, the $\log(m_b)$ dependence that arises when one b-quark is integrated out. Thus they can be regarded as an heuristic approach in order to obtain reasonable 5FNS predictions also for $t\bar{t}$+b-jet observables with a single resolved b jet. However, as discussed above, NLO+PS $t\bar{t}b\bar{b}$ predictions in the 5FNS are sensitive to the generation cuts in Eq. (B.1). In particular they depend on the unphysical parameters ξ_m and ξ_T, which have been set to one in the present study.[5] Moreover, one should keep in mind that $\log(m_b)$ terms beyond LO and finite terms of order m_b are not consistently included in the 5FNS.

[3] More precisely, MADGRAPH5_AMC@NLO employs the original formulation of MC@NLO matching [15, 55], while SHERPA+OPENLOOPS implements an alternative formulation [18,116] denoted as S-MC@NLO, which is characterised by an improved treatment of colour correlations but is otherwise equivalent to the method of [15, 55].

[4] To be precise, in the POWHEL simulation b-mass effects are neglected at the matrix element level but are taken into account in the parton shower.

[5] Variations of these parameters will allow to quantify up to which extent, in practice, predictions in conditions typically met in experimental analyses depend on the choice of these technical cuts.

Tools	Matching method	Shower	FNS	m_b [GeV]	Generation cuts
Sherpa 2.2.1+OpenLoops 1.2.3	S-mc@nlo	Sherpa	4FNS	4.75	fully inclusive
MadGraph5_aMC@NLO 2.3.2+Pythia8 2.1.0	MC@NLO	Pythia8	4FNS	4.75	fully inclusive
PowHel+Pythia8 2.1.0	Powheg	Pythia8	5FNS	0	$m_{bb} > 2m_b,\ p_{T,b} > m_b$

Table B.1: Employed tools, matching methods, parton showers, flavour number scheme (FNS) and generation cuts in the NLO+PS simulations of $t\bar{t}b\bar{b}$ production. The PowHel generator implements matrix elements with massless b-quarks but includes b-mass effects through Pythia.

B.3 Parton showers, PDF, and α_s

Full Monte Carlo simulations of $t\bar{t}b\bar{b}$ production involve hard $t\bar{t}b\bar{b}$ cross section at NLO, top-quark decays, parton showering, hadronisation, hadron decays, and the underlying event. The main source of theoretical uncertainty in this involved simulation framework is given by the mechanism that governs b-quark production in association with top-quark pairs. Thus, in order to obtain a sufficiently transparent picture of the nontrivial QCD dynamics of b-quark production, it was decided to reduce the complexity that results from the presence of the additional b-quarks that arise from top-quark decays via well-understood weak interactions. To this end, top quarks have been treated as stable particles in the simulations. Moreover, all NLO+PS simulations have been performed at the parton level, including only the perturbative phase of parton shower evolution, and neglecting hadronisation as well as any other non-perturbative aspect. The quantitative importance of hadronisation and the possible bias that can result from switching off hadronisation in the comparison of two $t\bar{t}$ + b-jet simulations based on different parton showers was assessed by comparing Sherpa 2.1 and Pythia 8.2 LO+PS simulations of $pp \to H$+jets (including b jets) at 14 TeV. Thanks to the color neutral nature of the Higgs boson, this process allows one to assess the impact of hadronisation by turning it on and off. The effects of hadronisation increase with decreasing jet transverse momenta. Thus they predominantly arise in the vicinity of the jet-p_T threshold. For the production of b-jets with $p_T > 25$ GeV and $|\eta| < 2.5$ they amount to about $-2\%(-4\%)$ per b jet in Pythia8 (Sherpa). This suggests that the bias that results from turning off hadronisation should be well below the typical NLO+PS uncertainties in $t\bar{t}b\bar{b}$ production.

In order to reduce uncontrolled sources of bias related to shower modelling in the comparison of NLO+PS simulations based on Sherpa and Pythia8,

those free parton-shower parameters related to the strong coupling have been chosen in a uniform way. Specifically, the rescaling factors x that are applied to the strong coupling terms $\alpha_S(x\,k_T^2)$ for each shower emission have been set to $x = 1$ both for initial- and final-state radiation. Furthermore the option of resumming subleading logarithms of Catani-Marchesini-Webber kind [51] was deactivated. Note that these choices neither correspond to the SHERPA default nor to the PYTHIA8 default settings. Moreover they are not meant to provide an optimal description of data. They are only aimed at a consistent comparison of the two showers, where simple parametric differences are avoided, and the remaining deviations can be attributed to intrinsic shower features, such as the different definition of the shower evolution variables.

Since parton-shower tunes and PDFs are intimately connected, it is not trivial to identify a common PDF set that is optimal for all parton showers. For the present study the NNPDF3.0 NLO PDF set was adopted, keeping in mind that this choice might bias the comparison of SHERPA against PYTHIA8. The specific PDF set was chosen according to the employed flavour number scheme (4FNS or 5FNS), while the value of $\alpha_s(M_Z)$ in NLO matrix elements and for the first shower emission was chosen consistently with the PDF. The same holds for the running of α_s, whose evolution is implemented at 2-loops both in matrix elements and parton showers. For subsequent shower emissions the 4FNS (5FNS) together with the corresponding value of $\alpha_s(M_Z)$ was used in SHERPA (PYTHIA8).

B.4 Input parameters and scale choices

To simulate $t\bar{t}b\bar{b}$ production at 13 TeV the input parameters $m_t = 172.5$ GeV, $m_b = 4.75$ GeV and $\alpha_s^{(5F)}(M_Z) = 0.118$ have been used together with NNPDF3.0 parton distributions at NLO, as discussed above.[6] The central values of the renormalisation and factorisation scales have been chosen as

$$\mu_{R,0} = \left(\prod_{i=t,\bar{t},b,\bar{b}} E_{T,i} \right)^{1/4}, \qquad \mu_{F,0} = \frac{H_T}{2} = \frac{1}{2} \sum_{i=t,\bar{t},b,\bar{b},j} E_{T,i}, \tag{B.2}$$

where $E_{T,i} = \sqrt{M_i^2 + p_{T,i}^2}$ denotes the transverse energy of top and bottom quarks, defined at parton level. Note that also extra parton emissions

[6] Note that the employed NNPDFs and related $\alpha_s(M_Z)$ value in the 4FNS are derived from variable-flavour-number NNPDFs with $\alpha_s^{(5F)}(M_Z) = 0.118$ via appropriate backward and forward evolution with five and four active flavours, respectively.

contribute to the total transverse energy H_T in Eq. (B.2). Theoretical uncertainties have been assessed by means of standard variations $\mu_R = \xi_R \mu_{R,0}$, $\mu_F = \xi_F \mu_{F,0}$ with $0.5 < \xi_R, \xi_F < 2$ and $0.5 < \xi_R/\xi_F < 2$.

The CKKW inspired renormalisation scale choice in Eq. (B.2) is based on [37] and takes into account the fact that top and bottom quarks are produced at widely different scales $E_{T,b} \ll E_{T,t}$. This turns out to improve the perturbative convergence as compared to a hard global scale of order m_t. In particular, in the 4FNS it was checked that using $\mu_R = H_T/2$ instead of $\mu_R = \mu_{R,0}$ increases the K-factor by 0.25 and reduces the NLO cross section by about 40%, which is only barely consistent with the level of uncertainty expected from factor-two scale variations. Moreover, computing LO and NLO cross sections using PDFs and α_s values at NLO throughout[7] yields K-factors around 2 with $\mu_R = \mu_{R,0}$ and about 0.25 higher with $\mu_R = H_T/2$. Thus both scale choices seem to be suboptimal, and in order to improve the convergence of the perturbative expansion, a scale even softer than Eq. (B.2) should be considered in the future. In any case a hard scale of type $\mu_R = H_T/2$ is not recommended.

In the context of the MC@NLO matching approach, where the resummation scale μ_Q, i.e. the parton shower starting scale, is a free parameter, it is natural to identify this scale with the factorisation scale. Thus $\mu_Q = \mu_{F,0} = H_T/2$ was used in the SHERPA+OPENLOOPS simulation. In the case of MADGRAPH5_AMC@NLO a different choice had to be adopted since only resummation scales of the form $\mu_Q = \xi\sqrt{\hat{s}}$ are supported, where the prefactor ξ is randomly distributed in the freely adjustable range $[\xi_{\min}, \xi_{\max}]$ with a distribution that is strongly peaked at $(\xi_{\min} + \xi_{\max})/2$ [27]. Comparing the $H_T/2$ and $\mu_Q = \xi\sqrt{\hat{s}}$ distributions it was observed that the respective peaks lie around 200 GeV and 400 GeV when the default MADGRAPH5_AMC@NLO settings $(\xi_{\max}, \xi_{\max}) = (0.1, 1)$ are used, i.e. the default μ_Q in MADGRAPH5_AMC@NLO is much harder.

Given that MC@NLO predictions for $t\bar{t}b\bar{b}$ production are quite sensitive to μ_Q, it was decided to lower the ξ upper bound to $\xi_{\max} = 0.25$, which brings the μ_Q reasonably close to $H_T/2$. We note that this choice is also supported by the study of an MADGRAPH5_AMC@NLO simulation of $Hb\bar{b}$ production in the 4FNS [123], where it was found that reducing $\xi_{\max}$ from 1 to 0.25 strongly improves the convergence of NLO+PS and NLO distributions at large transverse momenta.

In the POWHEG matching method, the resummation scale is not a freely

[7] With this approach K-factors are much less dependent on the employed PDF sets and reflect the convergence of the perturbative expansion in a more realistic way as compared to using LO inputs for the LO cross section.

adjustable parameter, since the first emission on top of $t\bar{t}b\bar{b}$ events is entirely described by matrix elements, and the corresponding transverse momentum scale sets the upper bound for subsequent shower emissions. Nevertheless, POWHEG simulations involve a parameter h_{damp} that separates the first-emission phase space into a singular region, where the first emission is resummed and corrected with a local K-factor, and a remnant region, where it is handled as at fixed-order NLO. Given the analogy with the separation of soft and hard events in the MC@NLO approach, and given that μ_Q represents the upper bound for emissions of soft events, it is natural to chose h_{damp} of the same order of μ_Q. Thus the choice $h_{\text{damp}} = H_T/2$ was adopted in the POWHEL simulation.

Variations of the resummation scale and of the h_{damp} parameter have not been considered in this study.

B.5 NLO+PS predictions for $t\bar{t} + b$-jets cross sections in b-jet bins

In the following we compare integrated and differential NLO+PS predictions for $t\bar{t} + b$-jets production with a certain minimum number of b jets, $n_b > N_b$. In particular we focus on the bins with $n_b \geq 1$ or $n_b \geq 2$, which are the most relevant ones for $t\bar{t}H(b\bar{b})$ analyses. For the jet definition the anti-k_T algorithm with $R = 0.4$ is adopted, and jets that involve one or more b-quark constituents are classified as b-jets. Note that also jets that result from collinear $g \to b\bar{b}$ splittings are handled as b jets. Moreover no requirement is imposed on the minimum transverse momentum of b quarks inside b jets. Events are categorised according to the number n_b of resolved b jets within the acceptance region,

$$p_{T,b} > 25 \text{ GeV}\,, \qquad |\eta_b| < 2.5\,. \tag{B.3}$$

Let us recall that top quarks are treated as stable particles, thus the two b quarks that arise from top decays as well as possible extra b quarks from the showering of top-decay products are not included in n_b. Apart from the requirement $n_b \geq N_b$ no additional cut will be applied.[8] In order to illustrate the importance of parton shower effects, the various NLO+PS predictions presented in the following are also compared to fixed-order NLO predictions.

[8] To be more precise, the SHERPA+OPENLOOPS and MADGRAPH5_AMC@NLO samples are fully inclusive, while in the case of POWHEL the technical cuts Eq. (B.1) are applied as discussed above.

Selection	Tool	σ_{NLO} [fb]	$\sigma_{\text{NLO+PS}}$ [fb]	$\sigma_{\text{NLO+PS}}/\sigma_{\text{NLO}}$
$n_b \geq 1$	SHERPA+OPENLOOPS	$12820^{+35\%}_{-28\%}$	$12939^{+30\%}_{-27\%}$	1.01
	MADGRAPH5_AMC@NLO		$13833^{+37\%}_{-29\%}$	1.08
	POWHEL		$10073^{+45\%}_{-29\%}$	0.79
$n_b \geq 2$	SHERPA+OPENLOOPS	$2268^{+30\%}_{-27\%}$	$2413^{+21\%}_{-24\%}$	1.06
	MADGRAPH5_AMC@NLO		$3192^{+38\%}_{-29\%}$	1.41
	POWHEL		$2570^{+35\%}_{-28\%}$	1.13

Table B.2: Fixed-order NLO and NLO+PS predictions for integrated $t\bar{t} + b$-jets cross sections at 13 TeV in bins with $n_b \geq 1$ and $n_b \geq 2$ b jets.

The latter are based on SHERPA+OPENLOOPS and are obviously independent of the employed parton shower and matching scheme.

All quoted theoretical uncertainties correspond to factor-two variations of the renormalisation and factorisation scales. In Figs. B.1–B.9 they are shown as bands, and, to improve readability, three different ratio plots are shown, where all results are normalised to one particular NLO QCD+PS prediction and the corresponding scale variation band is shown.

Results for the $t\bar{t} + b$-jets cross sections with $n_b \geq N_b$ b jets for various values of N_b are presented in Table B.2 and Figs. B.1. In the following we will refer to the results for $N_b = 1, 2, 3, 4$ as ttb, ttbb, tt + 3b and tt + 4b cross sections, respectively. For the ttb and ttbb cross sections, which are described at NLO accuracy, the various NLO+PS predictions turn out to be in decent mutual agreement. More precisely, ttb predictions based on the 4FNS (SHERPA+OPENLOOPS and MADGRAPH5_AMC@NLO) agree very well with each other and with fixed-order NLO, and the 5FNS ttb simulation (POWHEL) lies only 20% lower, despite that it was not designed to describe final states with a single b-jet (due to the generation cuts).

For the ttbb cross section one finds excellent agreement between fixed-order NLO, SHERPA+OPENLOOPS and POWHEL. This seems to suggest that this observable has little sensitivity to parton shower effects and to the choice of the flavour number scheme. However this interpretation is challenged by the fact that the MADGRAPH5_AMC@NLO ttbb result lies more than 30% above the other predictions. The only significant differences between MADGRAPH5_AMC@NLO and SHERPA+OPENLOOPS simulations lie in the employed parton showers and details of MC@NLO matching, thus the origin of the observed discrepancy is likely to lie in the choice of shower starting scale in MADGRAPH5_AMC@NLO combined with the higher intensity of QCD ra-

diation in PYTHIA8 with respect to SHERPA. This is confirmed by the further enhancement of the MADGRAPH5_AMC@NLO cross section in the bins with $n_b \geq 3$ and $n_b \geq 4$ b-jets (see Figs. B.1), where the additional b quarks arise from $g \to b\bar{b}$ parton-shower splittings, which results in a much stronger sensitivity to shower effects. Note that this kind of uncertainty for $N_b = 3, 4$ is not included in the quoted scale variations. In the SHERPA+OPENLOOPS simulation, the size of scale uncertainties and the difference between NLO and NLO+PS predictions are fairly similar to what observed at $\sqrt{s} = 8\,\text{TeV}$ in [37]. In particular, NLO+PS scale uncertainties range between 20–30% in all b-jet bins and are smaller as compared to the case of fixed-order NLO. Scale variations in MADGRAPH5_AMC@NLO and POWHEL tend to be larger and agree well with each other for $N_b = 2$, while POWHEL features a larger scale dependence in the other bins, especially for $N_b = 3, 4$. These various differences can be attributed to the employed flavour-number schemes and to technical aspects of the implementation of scale variations in the three different NLO+PS Monte Carlo tools.

B.6 ttb differential analysis

Various differential observables for an inclusive ttb analysis with $n_b \geq 1$ b-jets are presented in Figs. B.2–B.4. For all distributions that are inclusive with respect to extra light-jet emissions one observes a rather similar behaviour as for the ttb cross section, i.e. SHERPA+OPENLOOPS, MADGRAPH5_AMC@NLO and fixed-order NLO predictions agree well, while POWHEL lies about 20% lower. Only POWHEL features significant shape distortions with respect to fixed-order NLO in the region of low rapidity and/or low p_T for the leading top and bottom quarks and for the $t\bar{t}$ system (Figs. B.2–B.3). Observables that explicitly involve the first light-jet emission (Fig. B.4) turn out to behave differently. While for SHERPA+OPENLOOPS, POWHEL and fixed-order NLO there is mutual agreement within scale variations, the MADGRAPH5_AMC@NLO prediction turns out to lie up to 50% higher at $p_{T,j} \sim 50$ GeV. This enhancement of QCD radiation in MADGRAPH5_AMC@NLO +PYTHIA8 disappears at $p_{T,j} \sim 150$ GeV. It is most likely related to what was observed above in b-jet bin cross sections with $N_b \geq 2$.

B.7 ttbb differential analysis

Various differential observables for an inclusive ttbb analysis with $n_b \geq 2$ b-jets are presented in Figs. B.5–B.9 Observables that depend on the top-quark

and b-jet kinematics but are inclusive with respect to extra jet emission are presented in Figs. B.5–B.7. For all such distributions a fairly good agreement between SHERPA+OPENLOOPS, POWHEL and fixed-order NLO is observed, both at the level of shapes and normalisation. The most significant shape differences show up in the p_T of the 2nd b-jet and do not exceed 20%. In MADGRAPH5_AMC@NLO the matching to the PYTHIA8 shower increases the ttbb rates by about 35% with respect to SHERPA+OPENLOOPS, and turns out to have an non-trivial dependence on the top and b-jet kinematics. In particular it tends to enhance distributions in the regions with small top-quark and b-jet p_T and at large ΔR separation between the two b-jets.

For the distribution in the invariant mass of the b-jet pairs, which corresponds to the mass of the $H \to b\bar{b}$ candidate, it turns out that all NLO+PS results are in reasonably good mutual agreement. The results also confirm the presence of an NLO+PS distortion of the invariant-mass distribution, which was attributed to double-splitting effects in [37]. More precisely, in the vicinity of the Higgs resonance the NLO+PS enhancement w.r.t. NLO is close to 20% and thus less pronounced to what was observed in [37] at $\sqrt{s} = 8\,\text{TeV}$,[9] while the POWHEL and MADGRAPH5_AMC@NLO distributions feature an additional enhancement of about 10% and 35%, respectively, w.r.t. SHERPA+OPENLOOPS in the Higgs signal region.

Various observables that are directly sensitive to the emission of an additional jet are shown in Figs. B.8–B.9. Despite the intrinsic LO nature and stronger shower dependence of such distributions, SHERPA+OPENLOOPS and POWHEL remain in good agreement: the most important deviations, which show up in the p_T tail of the first light jet, do not exceed 40%. In contrast, the excess of MADGRAPH5_AMC@NLO w.r.t. the other predictions grows by about a factor two, reaching about 70% in average, and gives rise to more pronounced shape distortions as compared to the case of inclusive ttbb observables. Similarly as for the ttbb analysis, the enhancement is concentrated at light-jet momenta between 50–150 GeV, where it reaches up to 100%. A similarly strong increase shows up also in the region of central light-jet rapidity, as well as in angular and mass distributions that involve light and b-jets.

B.8 Summary and conclusions

In summary, we have presented a systematic study of Monte Carlo simulations of $pp \to t\bar{t} + b$-jets at 13 TeV that compares various NLO+PS pre-

[9] This can be due to the different collider energy and to the different scale choices in this study and in [37].

dictions based on different matching methods, parton showers and matching schemes. While the inclusion of b-mass effects is the only fully consistent way of describing inclusive $t\bar{t}+b$-jets production in terms of $t\bar{t}b\bar{b}$ matrix elements, the observed agreement between SHERPA+OPENLOOPS and POWHEL predictions indicates that also simulations with massless b-quarks and appropriate generation cuts provide predictions in agreement well within the scale uncertainties.

The various NLO+PS simulations considered in this study confirm that the invariant mass of the b-jet pair receives significant NLO+PS corrections that can reach 20-30% in the $H \to b\bar{b}$ signal region [37]. Based on standard variations of the renormalisation and factorisation scales, the expected accuracy of NLO predictions should be at the 25–35% level. However in various phase-space regions the differences between the various NLO+PS simulations tend to be larger.

In particular, some of the distributions generated with MADGRAPH5_AMC@NLO+PYTHIA8 have significantly different shapes, resulting in larger predictions for up to 100%, compared to the other NLO+PS simulations. These are probably related to the high intensity of the QCD radiation in PYTHIA8 and are quite sensitive to the choice of the shower starting scale in the MC@NLO matching framework. These findings should be regarded as a first step towards a thorough investigation of NLO matching and parton shower effects, including all relevant sources of uncertainty, in the Monte Carlo modelling of $t\bar{t}+b$-jets production. In the future also top-quark decays should be investigated.

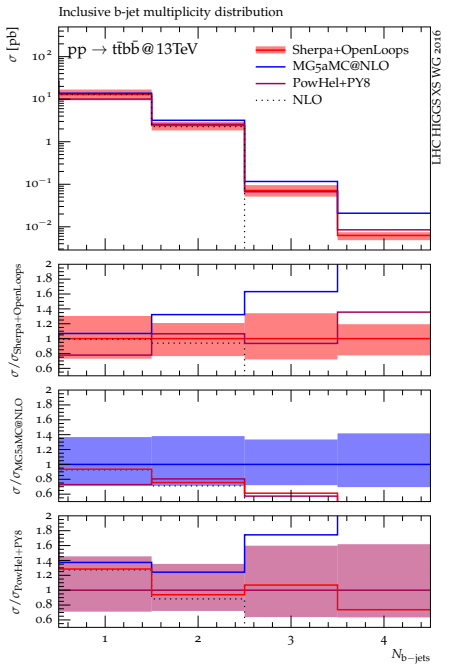


Figure B.1: Fixed-order NLO and NLO+PS predictions for integrated $t\bar{t}$+b-jets cross sections at 13 TeV in inclusive bins with $n_b > N_b$ b jets. Each ratio plot shows all results normalised to one particular NLO QCD+PS prediction and the corresponding scale variation band.

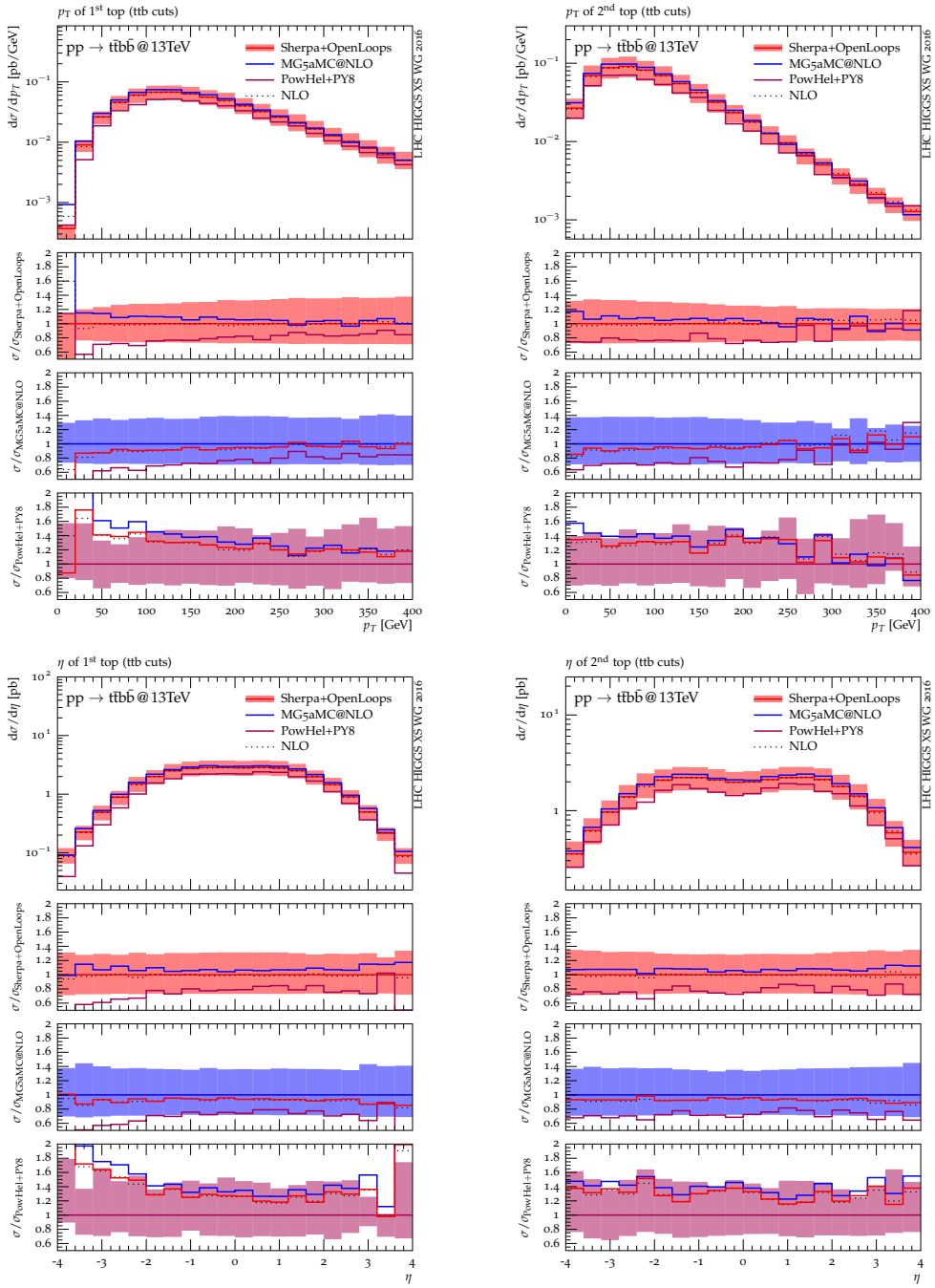


Figure B.2: Fixed-order NLO and NLO+PS top-quark distributions for $pp \to t\bar{t}+ \geq 1\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

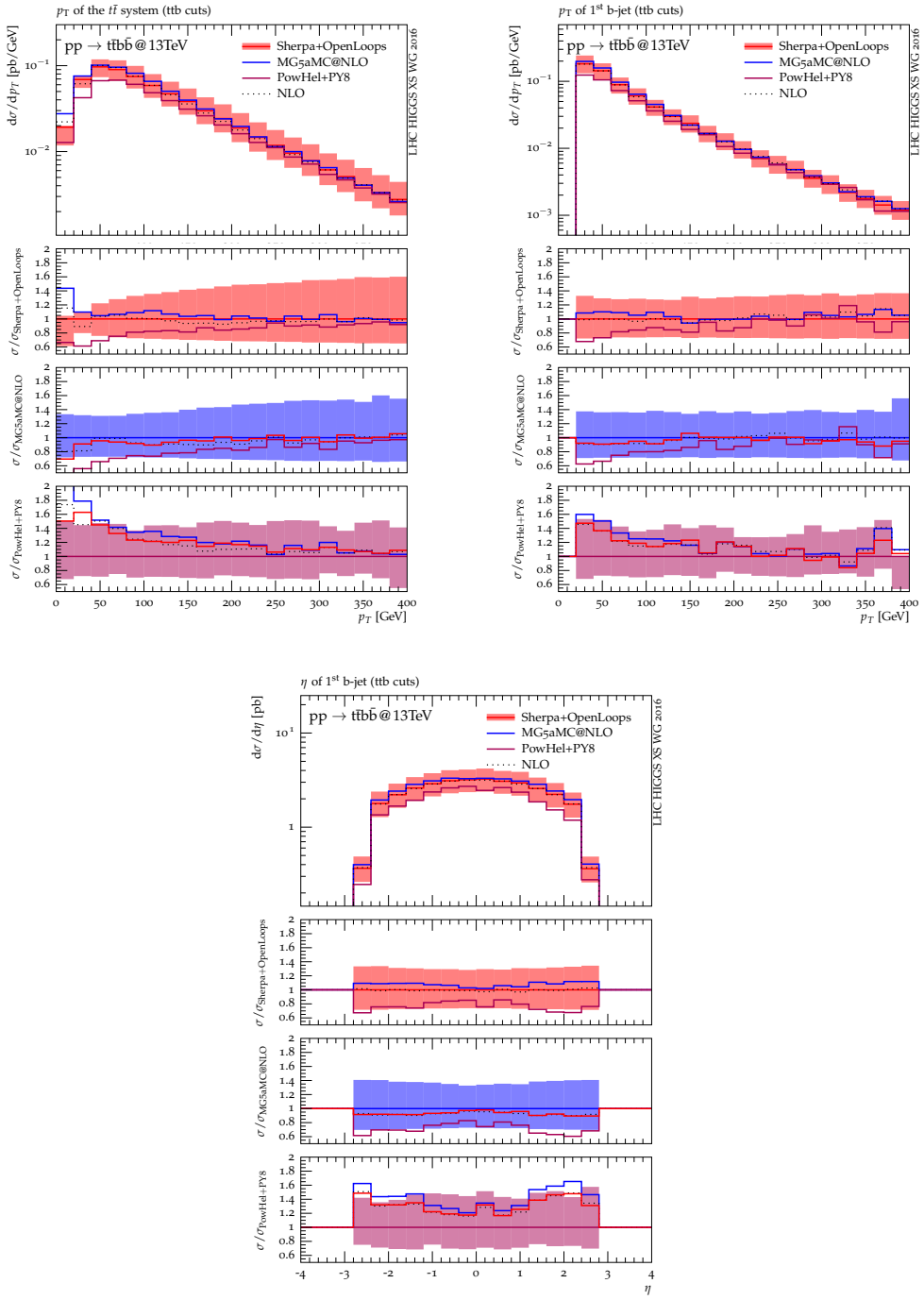


Figure B.3: Fixed-order NLO and NLO+PS top-quark and b-jet distributions for $pp \to t\bar{t}+ \geq 1\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

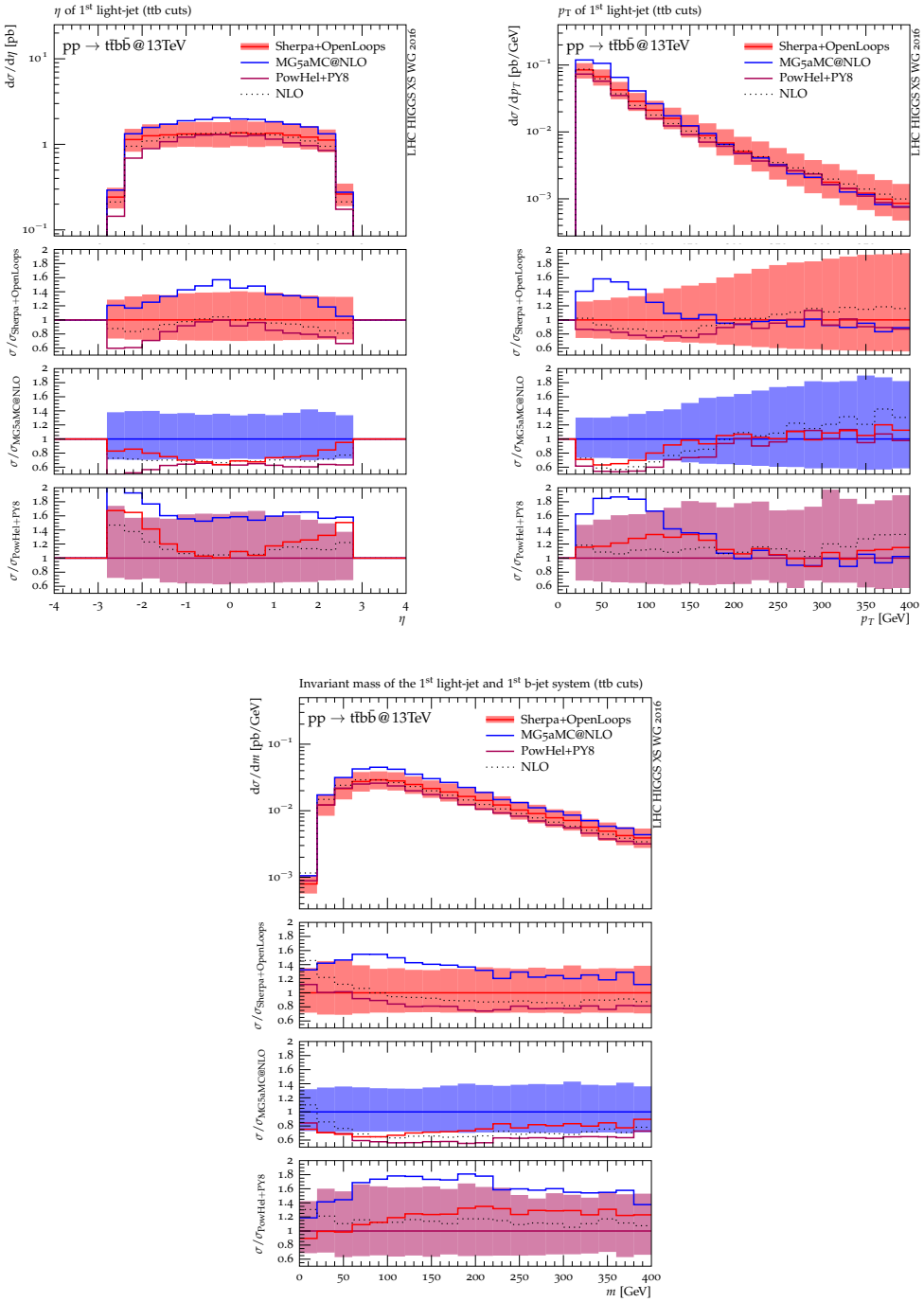


Figure B.4: Fixed-order NLO and NLO+PS light-jet distributions for $pp \to t\bar{t}+ \geq 1\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

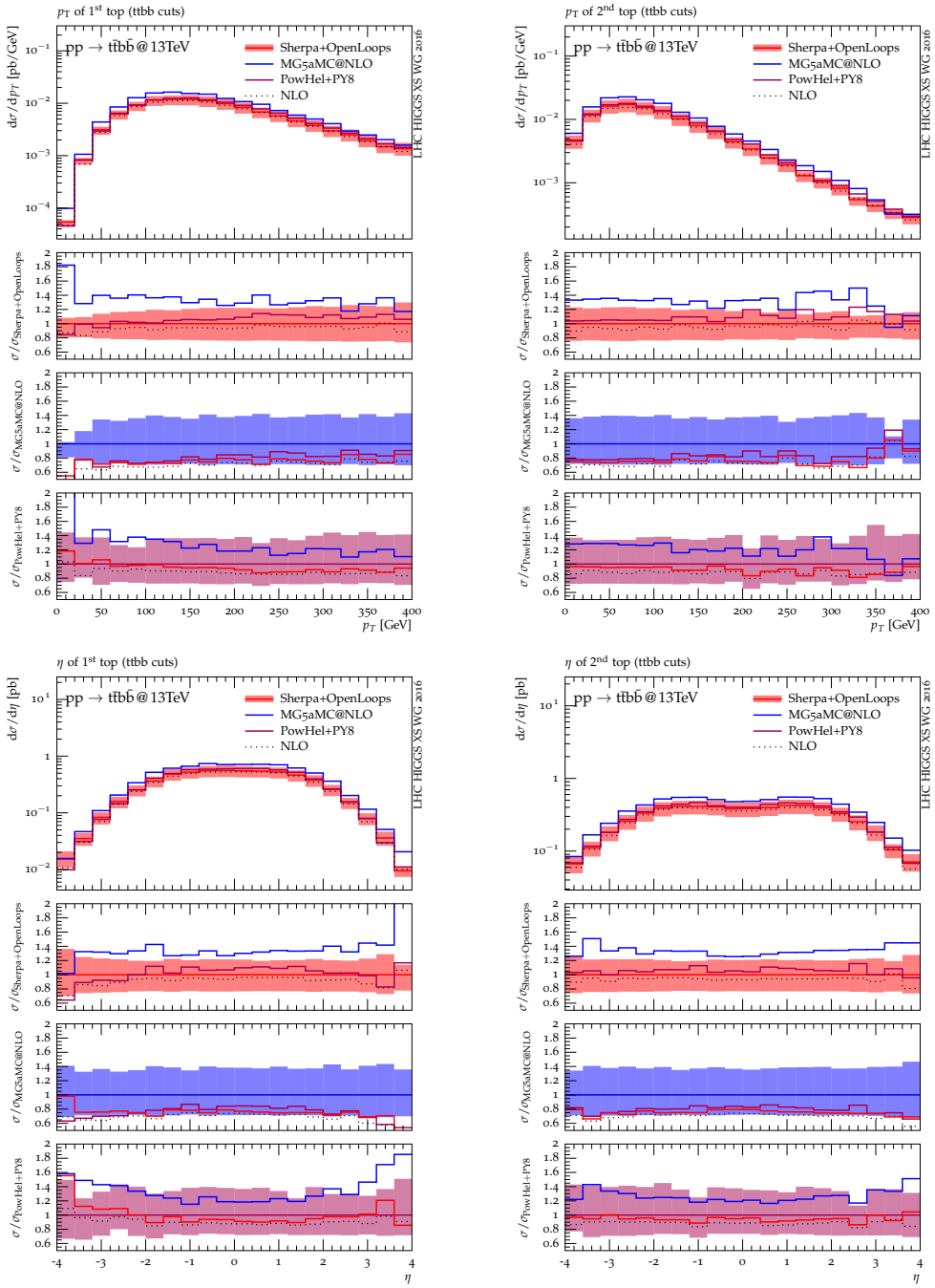


Figure B.5: Fixed-order NLO and NLO+PS top-quark distributions for $pp \to t\bar{t}+ \geq 2\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

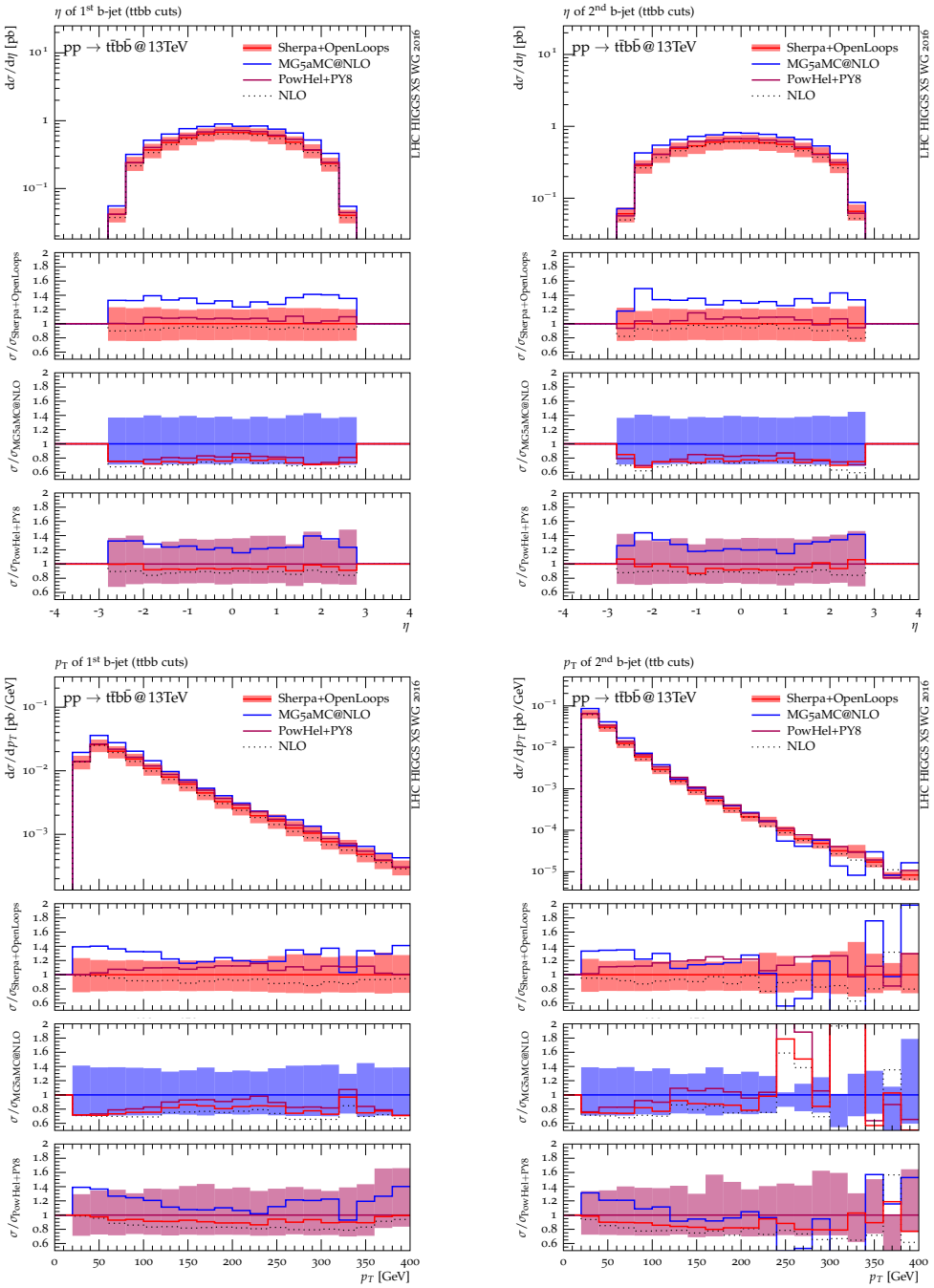


Figure B.6: Fixed-order NLO and NLO+PS b-jet distributions for $pp \to t\bar{t}+ \geq 2\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

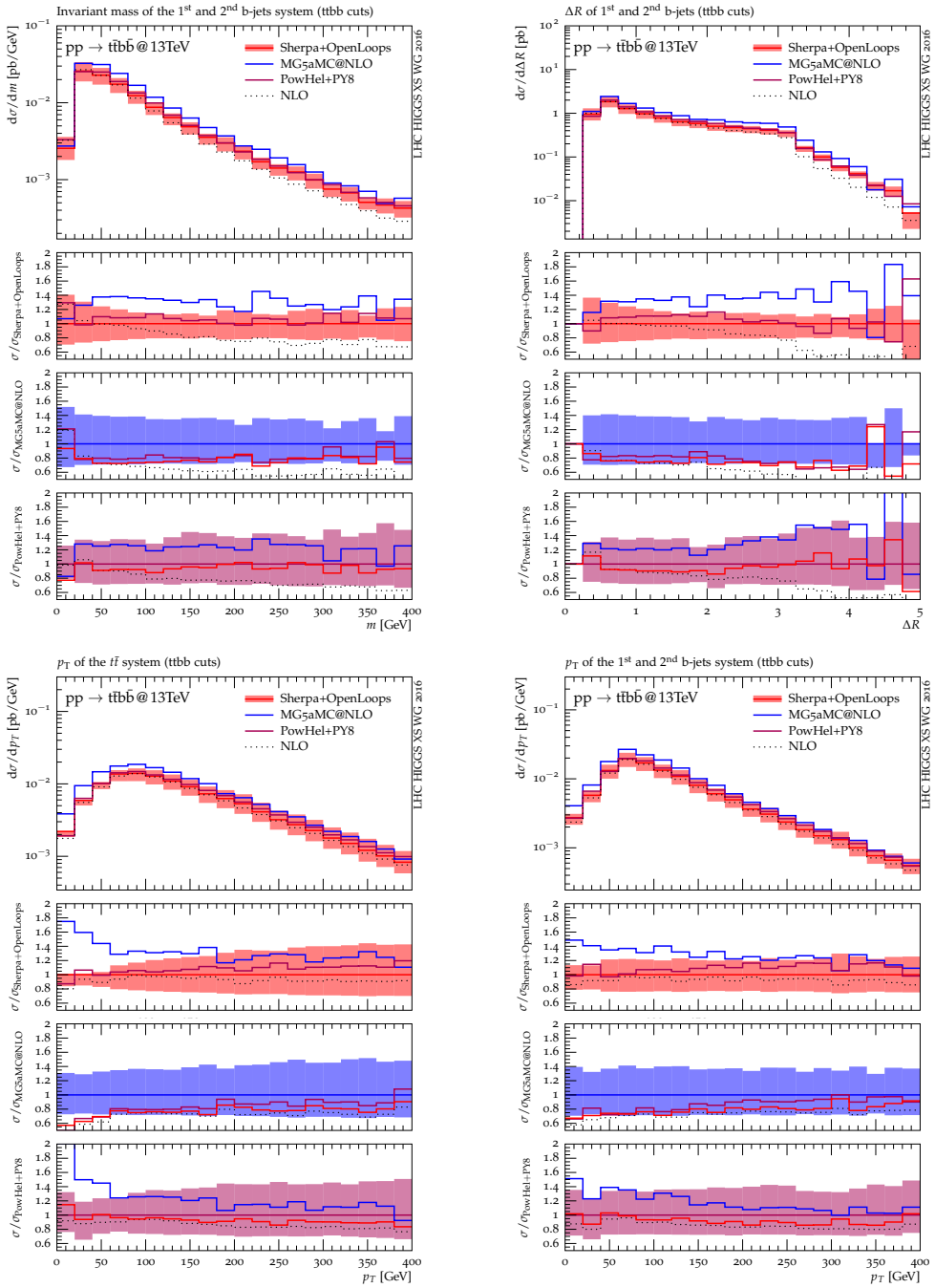


Figure B.7: (7) Fixed-order NLO and NLO+PS distributions of the $b\bar{b}$ and $t\bar{t}$ systems for $pp \to t\bar{t}+ \geq 2\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

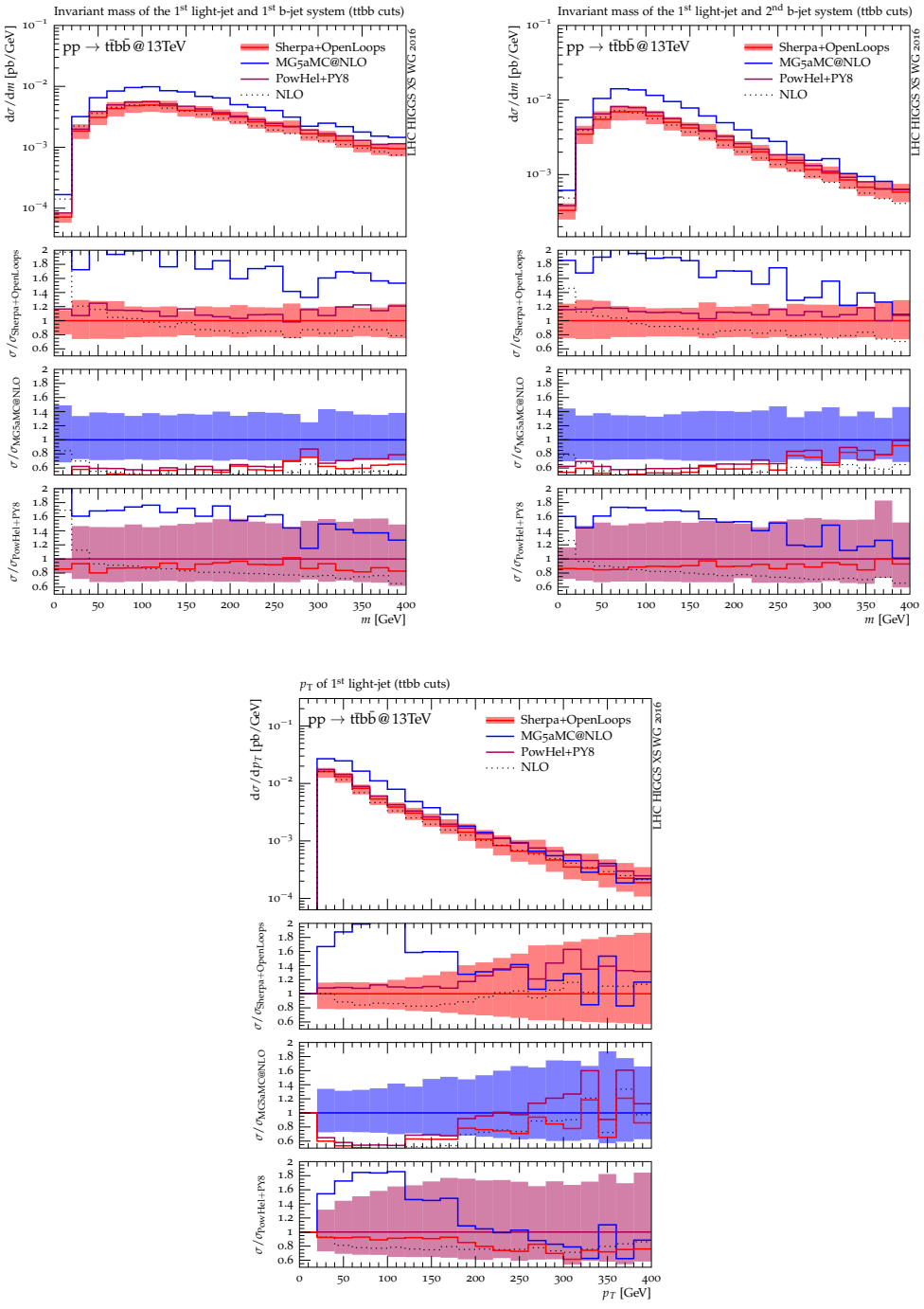


Figure B.8: Fixed-order NLO and NLO+PS light-jet mass and transverse momentum distributions for $pp \to t\bar{t}+ \geq 2\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

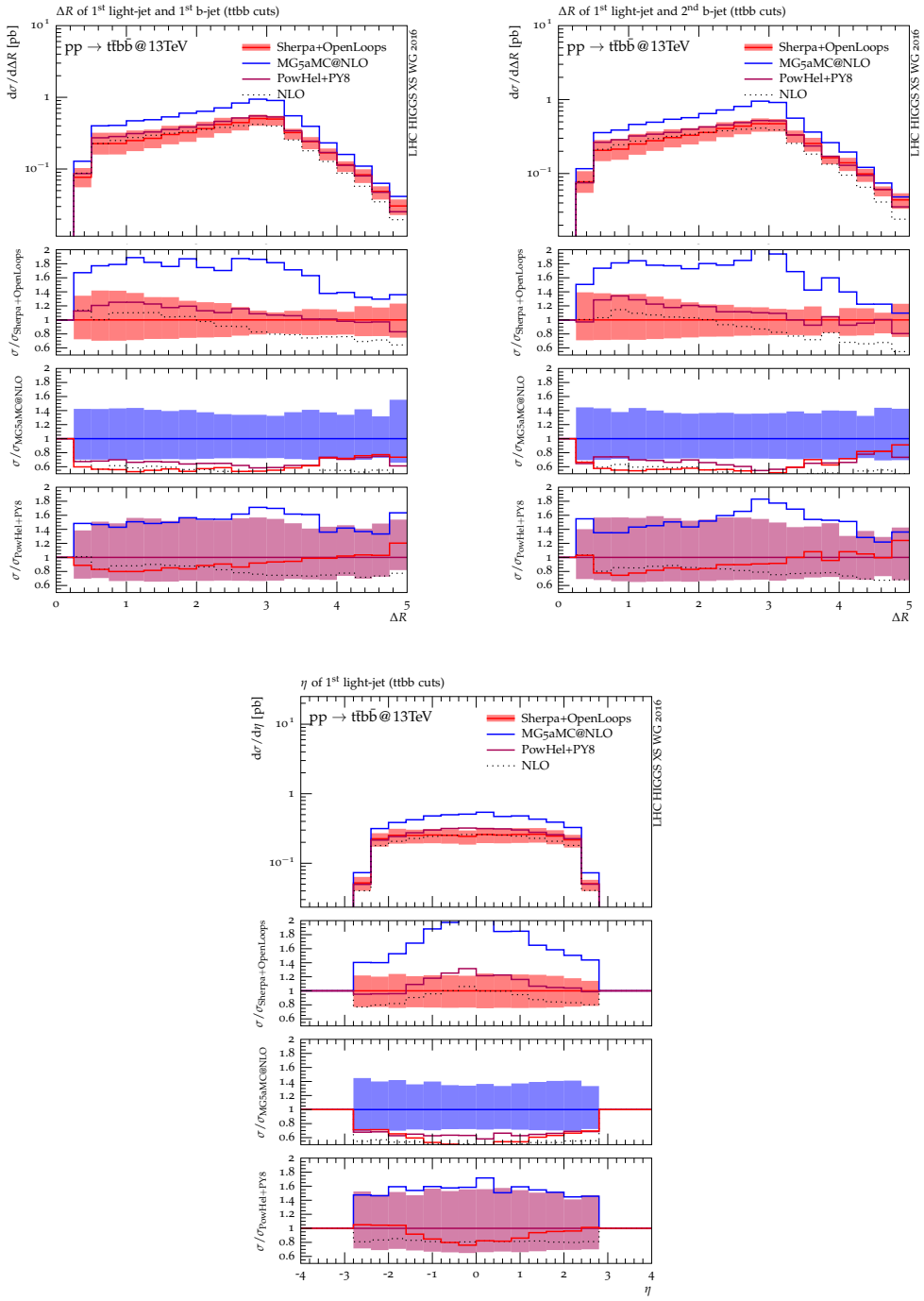


Figure B.9: Fixed-order NLO and NLO+PS light-jet angular distributions for $pp \to t\bar{t}+ \geq 2\,b$ jets at 13 TeV. Ratio plots like in Fig. B.1.

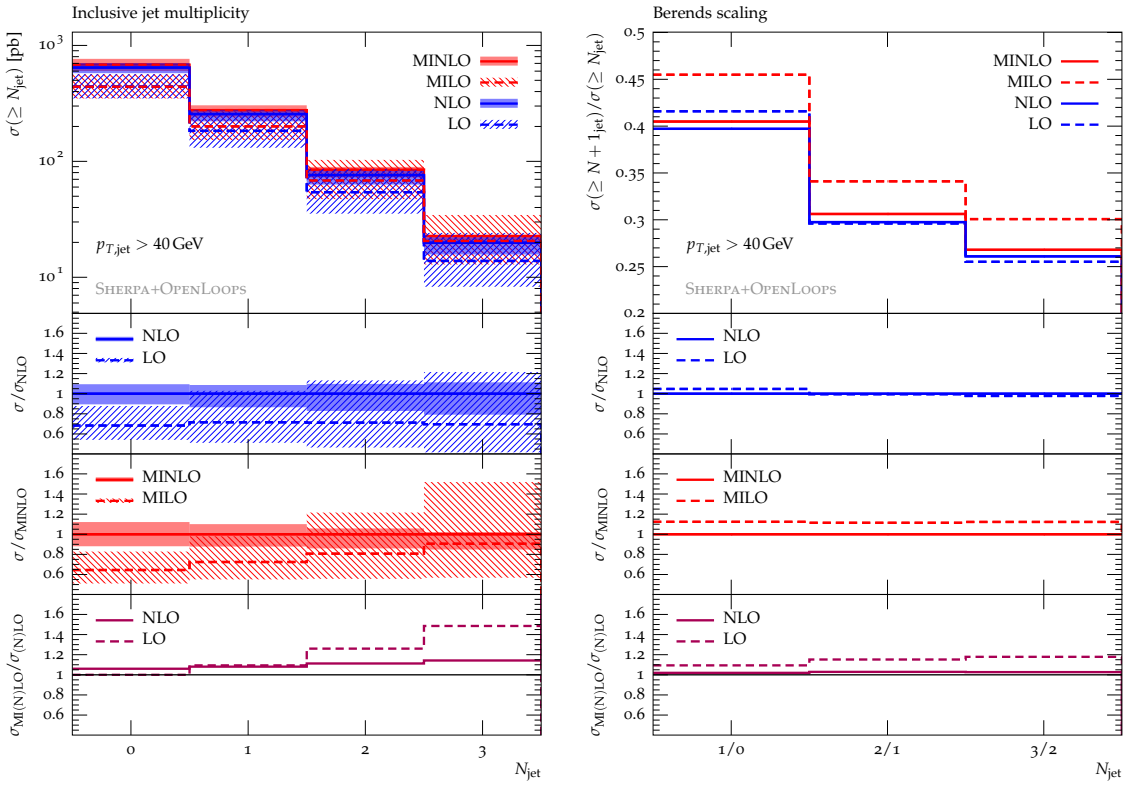
Appendix C

Further NLO predictions for $t\bar{t}$+multijet production

In this appendix we present additional NLO predictions for $t\bar{t}$+multijet production at 13 TeV, which have not been included in [39]. For a detailed description of the setup we refer to Section 3.2. We recall that jets are constructed by clustering light partons with the anti-k_t algorithm [65] with $R = 0.4$, and by default we select jets with pseudorapidity $|\eta_{\text{jet}}| < 2.5$ and a jet-p_{T} threshold of 25 GeV, unless stated otherwise.

In Figs. C.1-C.3 we show the $t\bar{t}+ \geq N$ jets inclusive cross sections and the ratio of $\sigma(t\bar{t}+ \geq N)$ over $\sigma(t\bar{t}+ \geq (N-1))$ using various $p_{\text{T,jet}}$ thresholds, $p_{\text{T,jet}} = 40, 60, 80$ GeV. Similarly as in Section 3.4, the top panel displays four predictions, stemming from fixed-order LO and NLO calculations, and from MiNLO computations at LO and NLO (labeled 'MILO' and 'MINLO'). The second panel shows the ratio between LO and NLO predictions at fixed order, while the third panel shows the ratio between MILO and MINLO predictions. The last panel shows the ratio between MINLO and NLO. The bands illustrate scale uncertainties estimated through independent factor-two rescaling of μ_{R} and μ_{F} excluding antipodal variations. These plots correspond to the results listed in Table 3.2 in Section 3.4. As also reported in Chapter 3, the good agreement between fixed-order NLO and MINLO results and the consistency of the observed NLO–MINLO differences with factor-two scale variations persist also for a range of other commonly used $p_{\text{T,jet}}$-thresholds [144].

In Fig. C.4 we show the pseudorapidity distributions of the hardest light jets in $t\bar{t}$ production plus one, two and three extra jets with $p_{\text{T,jet}} \geq 25$ GeV. These observables present quite flat K factors in both MINLO and NLO predictions, with a very good perturbative convergence. All the LO predictions exhibit a theoretical uncertainty of 30–40% for $N_{\text{jets}} \geq 1$ which raises

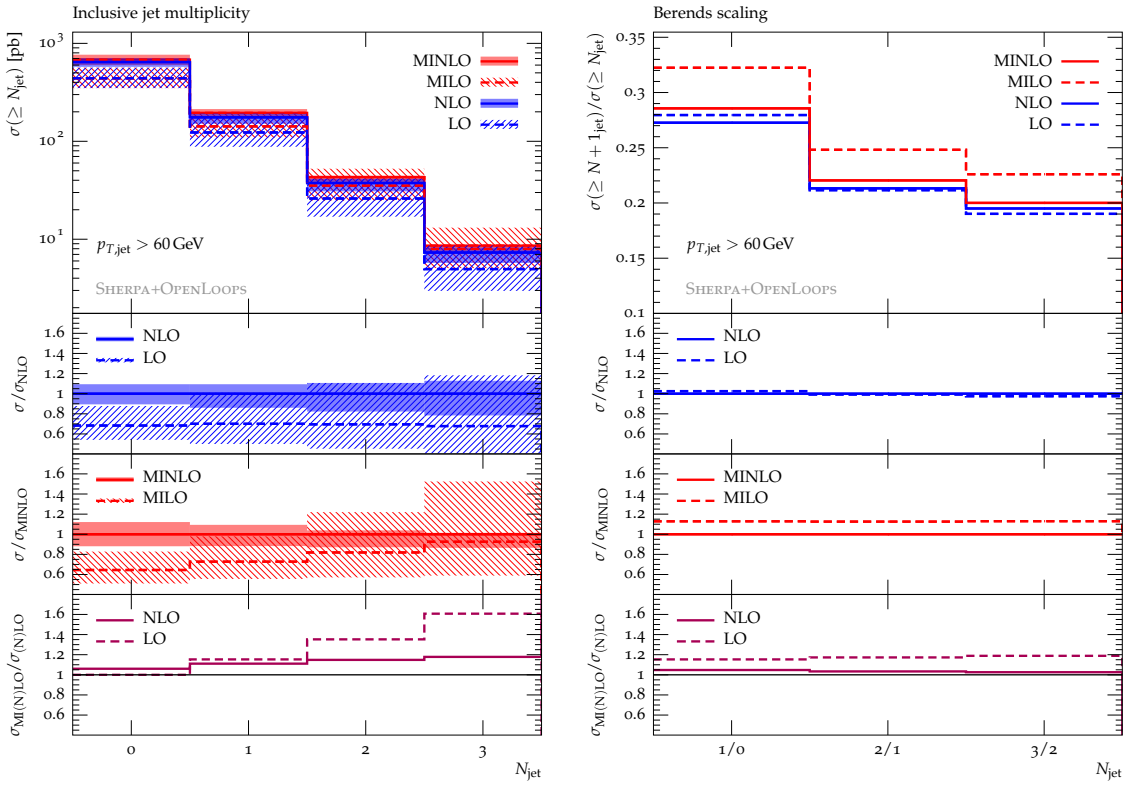


(a) (b)

Figure C.1: Inclusive $t\bar{t}$+multijet cross sections with a minimum number $N = 0, 1, 2, 3$ of jets (a) and ratios of $t\bar{t} + N$ jet over $t\bar{t} + (N-1)$ jet inclusive cross sections for $N = 1, 2, 3$ (b). Jets are defined with a p_T threshold of $p_{\text{T,jet}} \geq 40\,\text{GeV}$.

to 40–50% for $N_{\text{jets}} \geq 3$, whereas at NLO it shrinks to $\sim 10\%$ everywhere. Moreover, we can also observe that the agreement between the MINLO algorithm and the fixed order prediction with a scale $H_{\text{T}}/2$ is very good, but decreases with the number of jets. With $N_{\text{jets}} \geq 1$ we see that at LO and NLO the two simulations agree within 5%, whereas with $N_{\text{jets}} \geq 3$ the agreement is 30% and 10% at LO and NLO respectively, but anyway well within the theoretical uncertainties.

In Figs. C.5 and C.6 the azimuthal difference and the $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ difference between the hardest jets for $t\bar{t} + 2, 3$ jets are shown. The azimuthal separations between the hardest jets feature flat K factors. In passing from LO to NLO we observe a significant reduction of the theoretical uncertainty and tiny shape distortions everywhere. Again, we see a good agreement be-

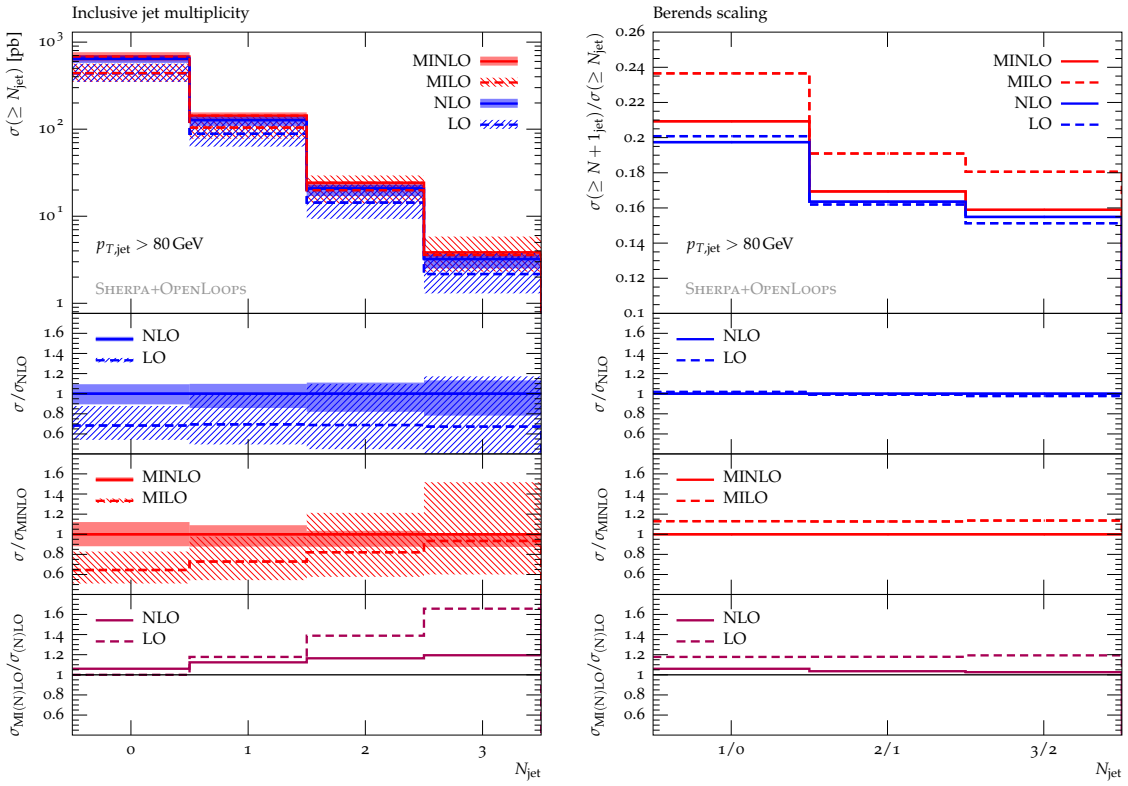


(a) (b)

Figure C.2: Inclusive $t\bar{t}$+multijet cross sections with a minimum number $N = 0, 1, 2, 3$ of jets (a) and ratios of $t\bar{t} + N$ jet over $t\bar{t} + (N-1)$ jet inclusive cross sections for $N = 1, 2, 3$ (b). Jets are defined with a p_T threshold of $p_{\text{T,jet}} \geq 60\,\text{GeV}$.

tween the MINLO algorithm and the fixed order, which passes from 30% at LO to 10% at NLO for $t\bar{t} + 3$ jet observables. Small shape distortions and small differences between the predictions are instead present in the ΔR separation between the hardest jets. Such distortions are mostly related to well separated jets. Here we observe that K factors are rather stable up to $\Delta R \sim 4$–5 in all the predictions, but the agreement between MINLO algorithm and the fixed order predictions decreases at higher values of R. For instance, the $\Delta R_{j_1,j_2}$ distribution for $t\bar{t} + 3$ jets exhibits a 50%–60% discrepancy at LO at $\Delta R \sim 5$. However, such distortions flatten at NLO in all the observables.

In summary, in this appendix, we have shown total inclusive cross sections with different jet $p_{\text{T,jet}}$ thresholds and angular observables which have not



(a) (b)

Figure C.3: Inclusive $t\bar{t}$+multijet cross sections with a minimum number $N = 0, 1, 2, 3$ of jets (a) and ratios of $t\bar{t} + N$ jet over $t\bar{t} + (N-1)$ jet inclusive cross sections for $N = 1, 2, 3$ (b). Jets are defined with a p_T threshold of $p_{\text{T,jet}} \geq 80\,\text{GeV}$.

been included in the original work [39]. All these observables confirm the good agreement between the MINLO and fixed order predictions claimed in the paper, and the differences turn out to be well consistent with standard scale variations.

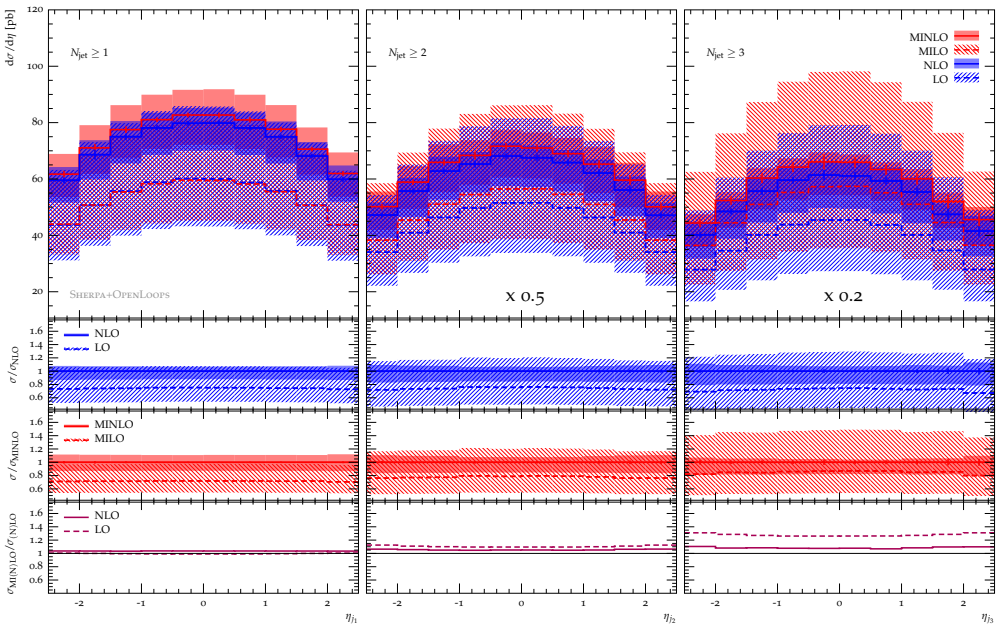


Figure C.4: Pseudorapidity distributions of the hardest light jets for $t\bar{t}$ + 1, 2, 3 jets with $p_{\text{T,jet}} \geq 25\,\text{GeV}$.

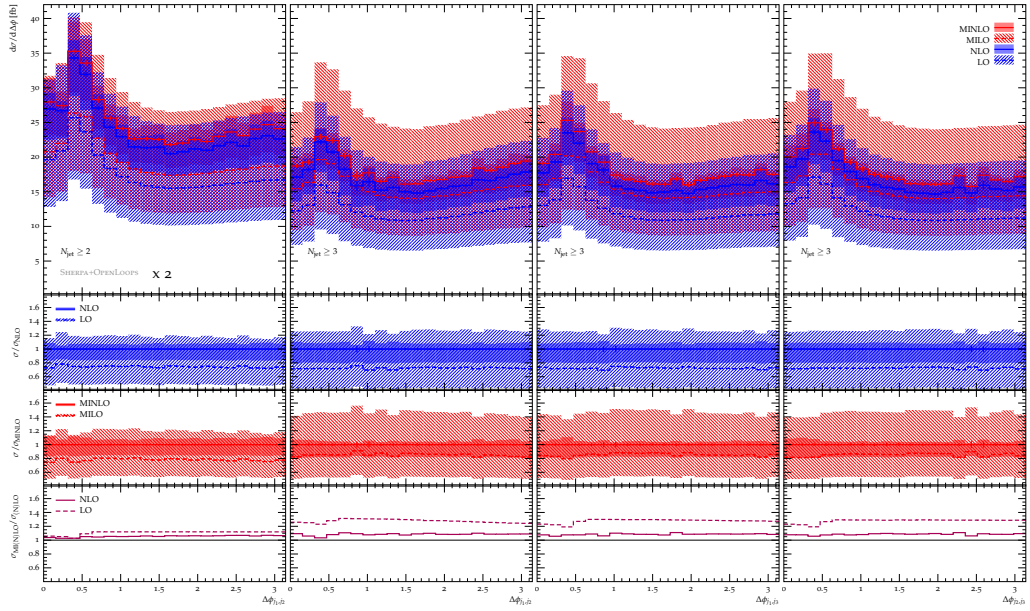


Figure C.5: Azimuthal difference distributions between the hardest light jets for $t\bar{t}+2,3$ jets with $p_{\mathrm{T,jet}} \geq 25\,\mathrm{GeV}$. The leftmost panel shows the observable $\Delta\phi_{j_1,j_2}$ for $t\bar{t}+2$ jets, whereas the other show respectively $\Delta\phi_{j_1,j_2}$, $\Delta\phi_{j_1,j_3}$ and $\Delta\phi_{j_2,j_3}$ for $t\bar{t}+3$ jets.

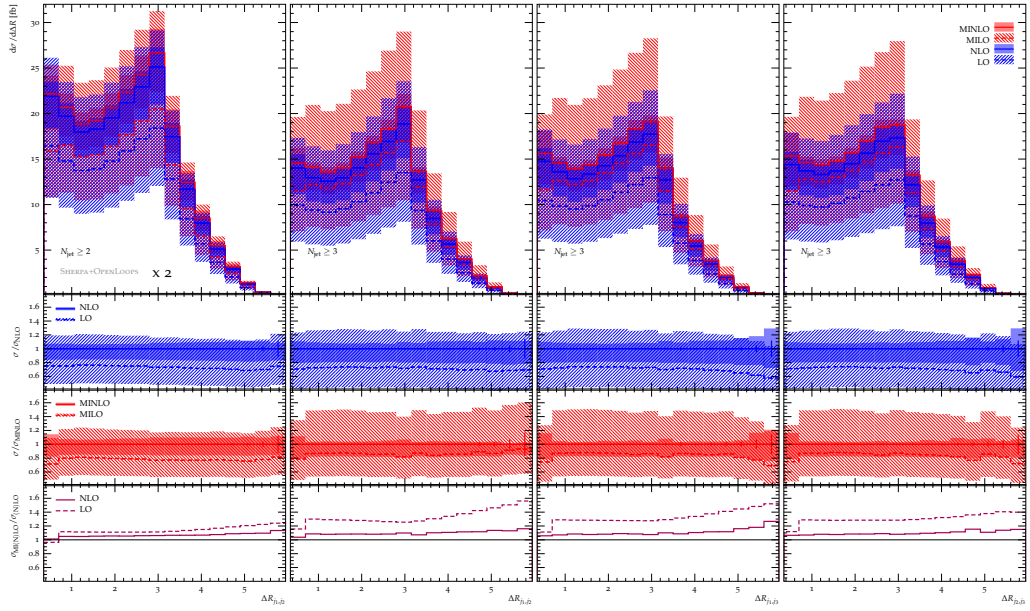


Figure C.6: ΔR distributions between the hardest light jets for $t\bar{t}+2,3$ jets with $p_{\mathrm{T,jet}} \geq 25\,\mathrm{GeV}$. The leftmost panel shows the observable $\Delta R_{j_1,j_2}$ for $t\bar{t}+2$ jets, whereas the other show respectively $\Delta R_{j_1,j_2}$, $\Delta R_{j_1,j_3}$ and $\Delta R_{j_2,j_3}$ for $t\bar{t}+3$ jets.

Bibliography

[1] ATLAS collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys.Lett.* **B716** (2012) 1–29, [1207.7214].

[2] CMS collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys.Lett.* **B716** (2012) 30–61, [1207.7235].

[3] S. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl.Phys.* **22** (1961) 579–588.

[4] S. Weinberg, *A Model of Leptons*, *Phys.Rev.Lett.* **19** (1967) 1264–1266.

[5] A. Salam, *Weak and Electromagnetic Interactions*, *Conf.Proc.* **C680519** (1968) 367–377.

[6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys.Rev.Lett.* **13** (1964) 321–323.

[7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys.Rev.Lett.* **13** (1964) 508–509.

[8] G. Guralnik, C. Hagen and T. Kibble, *Global Conservation Laws and Massless Particles*, *Phys.Rev.Lett.* **13** (1964) 585–587.

[9] ATLAS collaboration, T. A. collaboration, *Combination of the searches for Higgs boson production in association with top quarks in the $\gamma\gamma$, multilepton, and $b\bar{b}$ decay channels at $\sqrt{s}$=13 TeV with the ATLAS Detector*, .

[10] CMS collaboration, C. Collaboration, *Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13$ TeV*, .

[11] ATLAS collaboration, G. Aad et al., *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **05** (2016) 160, [1604.03812].

[12] CMS collaboration, V. Khachatryan et al., *Search for the associated production of the Higgs boson with a top-quark pair*, *JHEP* **09** (2014) 087, [1408.1682].

[13] ATLAS collaboration, G. Aad et al., *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C75** (2015) 349, [1503.05066].

[14] CMS collaboration, V. Khachatryan et al., *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method*, *Eur. Phys. J.* **C75** (2015) 251, [1502.02485].

[15] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029, [hep-ph/0204244].

[16] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[17] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [0709.2092].

[18] S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, [1111.1220].

[19] T. Gehrmann, S. Hoche, F. Krauss, M. Schonherr and F. Siegert, *NLO QCD matrix elements + parton showers in e^+e^- —¿ hadrons*, *JHEP* **01** (2013) 144, [1207.5031].

[20] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, *JHEP* **04** (2013) 027, [1207.5030].

[21] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063, [hep-ph/0109231].

[22] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, *JHEP* **08** (2002) 015, [hep-ph/0205283].

[23] M. L. Mangano, M. Moretti and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study*, *Nucl. Phys.* **B632** (2002) 343–362, [hep-ph/0108069].

[24] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061, [1209.6215].

[25] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [1002.2581].

[26] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, [0903.1219].

[27] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [1405.0301].

[28] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [0811.4622].

[29] G. Bevilacqua, M. Czakon, C. Papadopoulos and M. Worek, *Dominant QCD Backgrounds in Higgs Boson Analyses at the LHC: A Study of pp $\to$ t anti-t + 2 jets at Next-To-Leading Order*, *Phys. Rev. Lett.* **104** (2010) 162002, [1002.4009].

[30] G. Bevilacqua, M. Czakon, C. G. Papadopoulos and M. Worek, *Hadronic top-quark pair production in association with two jets at Next-to-Leading Order QCD*, *Phys. Rev.* **D84** (2011) 114017, [1108.2851].

[31] S. Höche, J. Huang, G. Luisoni, M. Schönherr and J. Winter, *Zero and one jet combined NLO analysis of the top quark forward-backward asymmetry*, *Phys.Rev.* **D88** (2013) 014040, [1306.2703].

[32] S. Hoeche, F. Krauss, P. Maierhoefer, S. Pozzorini, M. Schonherr and F. Siegert, *Next-to-leading order QCD predictions for top-quark pair production with up to two jets merged with a parton shower*, *Phys. Lett.* **B748** (2015) 74–78, [1402.6293].

[33] A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, *NLO QCD corrections to $pp \to t\bar{t}b\bar{b} + X$ at the LHC*, *Phys.Rev.Lett.* **103** (2009) 012002, [0905.0110].

[34] A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, *NLO QCD Corrections to Top Anti-Top Bottom Anti-Bottom Production at the LHC: 2. full hadronic results*, *JHEP* **03** (2010) 021, [1001.4006].

[35] G. Bevilacqua, M. Czakon, C. Papadopoulos, R. Pittau and M. Worek, *Assault on the NLO Wishlist:* $pp \to t\bar{t}b\bar{b}$, *JHEP* **09** (2009) 109, [0907.4723].

[36] A. Kardos and Z. Trócsányi, *Hadroproduction of t anti-t pair with a b anti-b pair using PowHel*, *J. Phys.* **G41** (2014) 075005, [1303.6291].

[37] F. Cascioli, P. Maierhoefer, N. Moretti, S. Pozzorini and F. Siegert, *NLO matching for ttbb production with massive b-quarks*, *Phys.Lett.* **B734** (2014) 210–214, [1309.5912].

[38] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: Multi-scale improved NLO*, *JHEP* **1210** (2012) 155, [1206.3572].

[39] S. Hche, P. Maierhoefer, N. Moretti, S. Pozzorini and F. Siegert, *Next-to-leading order QCD predictions for top-quark pair production with up to three jets*, 1607.06934.

[40] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization for Short Distance Hadron - Hadron Scattering*, *Nucl. Phys.* **B261** (1985) 104–142.

[41] J. C. Collins, D. E. Soper and G. F. Sterman, *Soft Gluons and Factorization*, *Nucl. Phys.* **B308** (1988) 833–856.

[42] T. Kinoshita, *Mass singularities of Feynman amplitudes*, *J. Math. Phys.* **3** (1962) 650–677.

[43] T. D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, *Phys. Rev.* **133** (1964) B1549–B1562.

[44] W. T. Giele and E. W. N. Glover, *Higher order corrections to jet cross-sections in e+ e- annihilation*, *Phys. Rev.* **D46** (1992) 1980–2010.

[45] W. T. Giele, E. W. N. Glover and D. A. Kosower, *Higher order corrections to jet cross-sections in hadron colliders*, *Nucl. Phys.* **B403** (1993) 633–670, [hep-ph/9302225].

[46] S. Catani and M. H. Seymour, *A General algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys.* **B485** (1997) 291–419, [hep-ph/9605323].

[47] S. Frixione, Z. Kunszt and A. Signer, *Three jet cross-sections to next-to-leading order*, *Nucl. Phys.* **B467** (1996) 399–442, [hep-ph/9512328].

[48] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145–233, [1101.2599].

[49] P. Nason and B. Webber, *Next-to-Leading-Order Event Generators*, *Ann. Rev. Nucl. Part. Sci.* **62** (2012) 187–213, [1202.1251].

[50] T. Sjostrand, *A Model for Initial State Parton Showers*, *Phys. Lett.* **B157** (1985) 321–325.

[51] S. Catani, B. R. Webber and G. Marchesini, *QCD coherent branching and semiinclusive processes at large x*, *Nucl. Phys.* **B349** (1991) 635–654.

[52] L. Lonnblad, *ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model*, *Comput. Phys. Commun.* **71** (1992) 15–31.

[53] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038, [0709.1027].

[54] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, *JHEP* **04** (2009) 002, [0812.0578].

[55] S. Frixione, P. Nason and B. R. Webber, *Matching NLO QCD and parton showers in heavy flavor production*, *JHEP* **08** (2003) 007, [hep-ph/0305252].

[56] S. Hoeche, S. Schumann and F. Siegert, *Hard photon production and matrix-element parton-shower merging*, *Phys.Rev.* **D81** (2010) 034026, [0912.3501].

[57] T. Carli, T. Gehrmann and S. Hoeche, *Hadronic final states in deep-inelastic scattering with Sherpa*, *Eur. Phys. J.* **C67** (2010) 73–97, [0912.3715].

[58] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Longitudinally-invariant $k_\perp$-clustering algorithms for hadron–hadron collisions*, *Nucl. Phys.* **B406** (1993) 187–224.

[59] F. Krauss and G. Rodrigo, *Resummed jet rates for e^+ e^- annihilation into massive quarks*, *Phys. Lett.* **B576** (2003) 135–142, [hep-ph/0303038].

[60] G. Rodrigo and F. Krauss, *Resummed jet rates for heavy quark production in e^+e^- annihilation*, *Eur. Phys. J.* **C33** (2004) 457–459, [hep-ph/0309325].

[61] S. Catani, S. Dittmaier and Z. Trocsanyi, *One loop singular behavior of QCD and SUSY QCD amplitudes with massive partons*, *Phys. Lett.* **B500** (2001) 149–160, [hep-ph/0011222].

[62] J. M. Campbell, R. K. Ellis, P. Nason and G. Zanderighi, *W and Z bosons in association with two jets using the POWHEG method*, *JHEP* **08** (2013) 005, [1303.5447].

[63] A. Buckley, J. Butterworth, L. Lönnblad, D. Grellscheid, H. Hoeth et al., *Rivet user manual*, *Comput.Phys.Commun.* **184** (2013) 2803–2819, [1003.0694].

[64] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.* **D93** (2016) 033006, [1506.07443].

[65] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].

[66] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [0710.3820].

[67] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J.* **C76** (2016) 196, [1512.01178].

[68] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001, [hep-ph/0206293].

[69] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **02** (2003) 027, [hep-ph/0208156].

[70] F. Krauss, R. Kuhn and G. Soff, *AMEGIC++ 1.0: A Matrix element generator in C++*, *JHEP* **02** (2002) 044, [hep-ph/0109036].

[71] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys.Rev.Lett.* **108** (2012) 111601, [1111.5206].

[72] S. Badger, B. Biedermann, P. Uwer and V. Yundin, *Numerical evaluation of virtual corrections to multi-jet production in massless QCD*, *Comput. Phys. Commun.* **184** (2013) 1981–1998, [1209.0100].

[73] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al., *Automated One-Loop Calculations with GoSam*, *Eur. Phys. J.* **C72** (2012) 1889, [1111.2034].

[74] F. Cascioli, S. Hche, F. Krauss, P. Maierhfer, S. Pozzorini and F. Siegert, *Precise Higgs-background predictions: merging NLO QCD and squared quark-loop corrections to four-lepton + 0,1 jet production*, *JHEP* **01** (2014) 046, [1309.0500].

[75] S. Hoeche, F. Krauss, S. Pozzorini, M. Schoenherr, J. M. Thompson and K. C. Zapp, *Triple vector boson production through Higgs-Strahlung with NLO multijet merging*, *Phys. Rev.* **D89** (2014) 093015, [1403.7516].

[76] S. Kallweit, J. M. Lindert, P. Maierhfer, S. Pozzorini and M. Schnherr, *NLO electroweak automation and precise predictions for W+multijet production at the LHC*, *JHEP* **04** (2015) 012, [1412.5157].

[77] N. Moretti, P. Petrov, S. Pozzorini and M. Spannowsky, *Measuring the signal strength in $t\bar{t}H$ with $H \to b\bar{b}$*, *Phys. Rev.* **D93** (2016) 014019, [1510.08468].

[78] S. Kallweit, J. M. Lindert, P. Maierhofer, S. Pozzorini and M. Schnherr, *NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging*, *JHEP* **04** (2016) 021, [1511.08692].

[79] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, [0808.3674].

[80] A. Ballestrero and E. Maina, *A New method for helicity calculations*, *Phys. Lett.* **B350** (1995) 225–233, [hep-ph/9403244].

[81] T. Gleisberg, F. Krauss, C. G. Papadopoulos, A. Schaelicke and S. Schumann, *Cross-sections for multiparticle final states at a linear collider*, *Eur. Phys. J.* **C34** (2004) 173–180, [hep-ph/0311273].

[82] K. Hagiwara, W. Kilian, F. Krauss, T. Ohl, T. Plehn, D. Rainwater et al., *Supersymmetry simulations with off-shell effects for CERN LHC and ILC*, *Phys. Rev.* **D73** (2006) 055005, [hep-ph/0512260].

[83] T. Gleisberg, F. Krauss, K. T. Matchev, A. Schalicke, S. Schumann and G. Soff, *Helicity formalism for spin-2 particles*, *JHEP* **09** (2003) 001, [hep-ph/0306182].

[84] C. Duhr, S. Hoeche and F. Maltoni, *Color-dressed recursive relations for multi-parton amplitudes*, *JHEP* **08** (2006) 062, [hep-ph/0607057].

[85] R. Kleiss and R. Pittau, *Weight optimization in multichannel Monte Carlo*, *Comput. Phys. Commun.* **83** (1994) 141–146, [hep-ph/9405257].

[86] F. A. Berends, R. Pittau and R. Kleiss, *All electroweak four fermion processes in electron - positron collisions*, *Nucl. Phys.* **B424** (1994) 308–342, [hep-ph/9404313].

[87] L. G. P., "VEGAS - An Adaptive Multi-dimensional Integration Program."

[88] T. Gleisberg and F. Krauss, *Automating dipole subtraction for QCD NLO calculations*, *Eur. Phys. J.* **C53** (2008) 501–523, [0709.2881].

[89] Z. Nagy and D. E. Soper, *Matching parton showers to NLO computations*, *JHEP* **10** (2005) 024, [hep-ph/0503053].

[90] Z. Nagy and D. E. Soper, *A New parton shower algorithm: Shower evolution, matching at leading and next-to-leading order level*, in *Proceedings, Ringberg Workshop on New Trends in HERA Physics 2005*, pp. 101–123, 2006. hep-ph/0601021.

[91] G. Altarelli, N. Amapane, J. Andersen, V. Andreev, M. Arneodo, V. Avati et al., eds., *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*, 2005.

[92] T. Sjostrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, *Phys. Rev.* **D36** (1987) 2019.

[93] J.-C. Winter, F. Krauss and G. Soff, *A Modified cluster hadronization model*, *Eur. Phys. J.* **C36** (2004) 381–395, [hep-ph/0311085].

[94] F. Cascioli, S. Hche, F. Krauss, P. Maierhfer, S. Pozzorini and F. Siegert, *Automatic one-loop calculations with Sherpa+OpenLoops*, *J. Phys. Conf. Ser.* **523** (2014) 012058.

[95] G. Ossola, C. G. Papadopoulos and R. Pittau, *Reducing full one-loop amplitudes to scalar integrals at the integrand level*, *Nucl. Phys.* **B763** (2007) 147–169, [hep-ph/0609007].

[96] Fabio Cascioli and Jonas Lindert and Philipp Maierhöfer and Stefano Pozzorini, *The OpenLoops one-loop generator, publicly available at* *http://openloops.hepforge.org*.

[97] T. Hahn, *Generating Feynman diagrams and amplitudes with FeynArts 3*, *Comput. Phys. Commun.* **140** (2001) 418–431, [hep-ph/0012260].

[98] A. Denner and S. Dittmaier, *Reduction of one loop tensor five point integrals*, *Nucl. Phys.* **B658** (2003) 175–202, [hep-ph/0212259].

[99] A. Denner and S. Dittmaier, *Reduction schemes for one-loop tensor integrals*, *Nucl. Phys.* **B734** (2006) 62–115, [hep-ph/0509141].

[100] A. Denner and S. Dittmaier, *Scalar one-loop 4-point integrals*, *Nucl. Phys.* **B844** (2011) 199–242, [1005.2076].

[101] A. Denner, S. Dittmaier and L. Hofer, *Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations*, 1604.06792.

[102] G. Ossola, C. G. Papadopoulos and R. Pittau, *CutTools: A Program implementing the OPP reduction method to compute one-loop amplitudes*, *JHEP* **0803** (2008) 042, [0711.3596].

[103] A. van Hameren, *OneLOop: For the evaluation of one-loop scalar functions*, *Comput.Phys.Commun.* **182** (2011) 2427–2438, [1007.4716].

[104] M. Grazzini, S. Kallweit, D. Rathlev and A. Torre, *$Z\gamma$ production at hadron colliders in NNLO QCD*, *Phys. Lett.* **B731** (2014) 204–207, [1309.7000].

[105] F. Cascioli, S. Kallweit, P. Maierhfer and S. Pozzorini, *A unified NLO description of top-pair and associated Wt production*, *Eur. Phys. J.* **C74** (2014) 2783, [1312.0546].

[106] P. Maierhfer and A. Papaefstathiou, *Higgs Boson pair production merged to one jet*, *JHEP* **03** (2014) 126, [1401.0007].

[107] G. Abelof, A. Gehrmann-De Ridder, P. Maierhofer and S. Pozzorini, *NNLO QCD subtraction for top-antitop production in the $q\bar{q}$ channel*, *JHEP* **08** (2014) 035, [1404.6493].

[108] F. Cascioli, T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhfer, A. von Manteuffel et al., *ZZ production at hadron colliders in NNLO QCD*, *Phys. Lett.* **B735** (2014) 311–313, [1405.2219].

[109] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhfer, A. von Manteuffel, S. Pozzorini et al., *W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD*, *Phys. Rev. Lett.* **113** (2014) 212001, [1408.5243].

[110] CMS Collaboration collaboration, S. Chatrchyan et al., *Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC*, *JHEP* **1305** (2013) 145, [1303.0763].

[111] The ATLAS Collaboration, *ATLAS-CONF-2012-135*, 2012.

[112] M. Worek, *On the next-to-leading order QCD K-factor for top $t\bar{t}b\bar{b}$ production at the TeVatron*, *JHEP* **1202** (2012) 043, [1112.4325].

[113] F. Cascioli, P. Maierhoefer and S. Pozzorini.

[114] A. Denner, D. Dittmaier and L. Hofer.

[115] S. Catani, S. Dittmaier, M. H. Seymour and Z. Trocsanyi, *The Dipole formalism for next-to-leading order QCD calculations with massive partons*, *Nucl. Phys.* **B627** (2002) 189–265, [hep-ph/0201036].

[116] S. Hoeche and M. Schonherr, *Uncertainties in next-to-leading order plus parton shower matched simulations of inclusive jet and dijet production*, *Phys.Rev.* **D86** (2012) 094042, [1208.2815].

[117] A. Martin et al., *Parton distributions for the LHC*, *Eur.Phys.J.* **C63** (2009) 189–285, [0901.0002].

[118] The LHC Higgs Cross Section Working Group collaboration, S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 1307.1347.

[119] F. Maltoni, G. Ridolfi and M. Ubiali, *b-initiated processes at the LHC: a reappraisal*, *JHEP* **1207** (2012) 022, [1203.6393].

[120] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin et al., *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010) 074024, [1007.2241].

[121] N. Moretti and S. Pozzorini, *Monte Carlo samples for $t\bar{t}b\bar{b}$ production at 8 TeV, publicly available at* ***http://www.physik.uzh.ch/data/ttHsim/TTBBV1***.

[122] LHC Higgs Cross Section Working Group, *This study will appear in the fourth Handbook of LHC Higgs Cross Sections (in preparation) and is anticipated in Appendix B.*

[123] M. Wiesemann, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni and P. Torrielli, *Higgs production in association with bottom quarks*, *JHEP* **02** (2015) 132, [1409.5301].

[124] M. Czakon, P. Fiedler and A. Mitov, *The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$*, *Phys.Rev.Lett.* **110** (2013) 252004, [1303.6254].

[125] M. Czakon, D. Heymes and A. Mitov, *High-precision differential predictions for top-quark pairs at the LHC*, *Phys. Rev. Lett.* **116** (2016) 082003, [1511.00549].

[126] S. Dittmaier, P. Uwer and S. Weinzierl, *NLO QCD corrections to t anti-t + jet production at hadron colliders*, *Phys.Rev.Lett.* **98** (2007) 262002, [hep-ph/0703120].

[127] S. Frixione, P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, [0707.3088].

[128] A. Kardos, C. Papadopoulos and Z. Trocsanyi, *Top quark pair production in association with a jet with NLO parton showering*, *Phys.Lett.* **B705** (2011) 76–81, [1101.2672].

[129] S. Alioli, S.-O. Moch and P. Uwer, *Hadronic top-quark pair-production with one jet and parton showering*, *JHEP* **01** (2012) 137, [1110.5251].

[130] M. Czakon, H. B. Hartanto, M. Kraus and M. Worek, *Matching the Nagy-Soper parton shower at next-to-leading order*, *JHEP* **06** (2015) 033, [1502.00925].

[131] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres-Cordero, D. Forde, T. Gleisberg et al., *Precise Predictions for W + 4-Jet Production at the Large Hadron Collider*, *Phys. Rev. Lett.* **106** (2011) 092001, [1009.2338].

[132] H. Ita, Z. Bern, L. J. Dixon, F. Febres-Cordero, D. A. Kosower and D. Maître, *Precise Predictions for Z + 4 Jets at Hadron Colliders*, *Phys.Rev.* **D85** (2012) 031501, [1108.2229].

[133] Z. Bern, L. Dixon, F. Febres Cordero, S. Höche, H. Ita, D. A. Kosower et al., *Next-to-Leading Order W + 5-Jet Production at the LHC*, *Phys.Rev.* **D88** (2013) 014025, [1304.1253].

[134] S. Badger, B. Biedermann, P. Uwer and V. Yundin, *Next-to-leading order QCD corrections to five jet production at the LHC*, *Phys.Rev.* **D89** (2013) 034019, [1309.6585].

[135] S. Badger, A. Guffanti and V. Yundin, *Next-to-leading order QCD corrections to di-photon production in association with up to three jets at the Large Hadron Collider*, *JHEP* **1403** (2014) 122, [1312.5927].

[136] A. Denner and R. Feger, *NLO QCD corrections to off-shell top-antitop production with leptonic decays in association with a Higgs boson at the LHC*, *JHEP* **11** (2015) 209, [1506.07448].

[137] G. Bevilacqua, H. B. Hartanto, M. Kraus and M. Worek, *Top Quark Pair Production in Association with a Jet with Next-to-Leading-Order QCD Off-Shell Effects at the Large Hadron Collider*, *Phys. Rev. Lett.* **116** (2016) 052003, [1509.09242].

[138] A. Denner and M. Pellen, *NLO electroweak corrections to off-shell top-antitop production with leptonic decays at the LHC*, 1607.05571.

[139] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres-Cordero, D. Forde, T. Gleisberg et al., *Next-to-leading order QCD predictions for W+3-Jet distributions at hadron colliders*, *Phys. Rev.* **D80** (2009) 074036, [0907.1984].

[140] T. Gleisberg, S. Höche, F. Krauss, A. Schälicke, S. Schumann and J. Winter, *SHERPA1.α, a proof-of-concept version*, *JHEP* **02** (2004) 056, [hep-ph/0311263].

[141] Z. Bern et al., *Ntuples for NLO Events at Hadron Colliders*, *Comput.Phys.Commun.* **185** (2014) 1443–1460, [1310.7439].

[142] D. Amati, A. Bassetto, M. Ciafaloni, G. Marchesini and G. Veneziano, *A treatment of hard processes sensitive to the infrared structure of QCD*, *Nucl. Phys.* **B173** (1980) 429.

[143] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock and B. R. Webber, *New clustering algorithm for multijet cross sections in e^+e^- annihilation*, *Phys. Lett.* **B269** (1991) 432–438.

[144] ATLAS collaboration, *Measurement of jets produced in top quark events using the di-lepton final state with 2 b-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, .

[145] Z. Bern, L. Dixon, F. Febres Cordero, S. Höche, H. Ita, D. Kosower et al., *Extrapolating W-Associated Jet-Production Ratios at the LHC*, *Phys. Rev.* **D92** (2015) 014008, [1412.4775].

[146] BlackHat collaboration, Z. Bern, L. J. Dixon, F. Febres Cordero, S. Höche, H. Ita, D. A. Kosower et al., *Universality in W+Multijet Production*, 1407.6564.

[147] M. Rubin, G. P. Salam and S. Sapeta, *Giant QCD K-factors beyond NLO*, *JHEP* **09** (2010) 084, [1006.2144].

[148] M. V. Garzelli, A. Kardos and Z. Trócsányi, *t tbar b bbar hadroproduction at NLO accuracy matched with parton shower*, *PoS* **EPS-HEP2013** (2013) 253.

[149] M. V. Garzelli, A. Kardos and Z. Trócsányi, *Hadroproduction of $\mathrm{t\bar{t}b\bar{b}}$ final states at LHC: predictions at NLO accuracy matched with Parton Shower*, *JHEP* **03** (2015) 083, [1408.0266].

[150] L. Lönnblad and S. Prestel, *Merging Multi-leg NLO Matrix Elements with Parton Showers*, *JHEP* **03** (2013) 166, [1211.7278].

[151] V. Hirschi, R. Frederix, S. Frixione, M. V. Garzelli, F. Maltoni and R. Pittau, *Automation of one-loop QCD corrections*, *JHEP* **05** (2011) 044, [1103.0621].

[152] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *European Physical Journal* **74** (2014) 3024, [1404.5630].

[153] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [1410.3012].

[154] G. Bevilacqua et al., *HELAC-NLO*, 1110.1499.